

Received July 22, 2020, accepted August 2, 2020, date of publication August 14, 2020, date of current version September 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016893

# Coping With Multiple Visual Motion Cues Under Extremely Constrained Computation Power of Micro Autonomous Robots

CHENG HU<sup>1,2</sup>, (Member, IEEE), CAIHUA XIONG<sup>3</sup>, (Member, IEEE), JIGEN PENG<sup>4</sup>,  
AND SHIGANG YUE<sup>1,2</sup>, (Senior Member, IEEE)

<sup>1</sup>Machine Life and Artificial Intelligence Research Center, School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China

<sup>2</sup>Lincoln Centre for Autonomous Systems, University of Lincoln, Lincoln LN6 7TS, U.K.

<sup>3</sup>School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 432700, China

<sup>4</sup>School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China

Corresponding author: Shigang Yue (syue@lincoln.ac.uk)

This work was supported in part by the EU HORIZON 2020 Project Spatial-temporal information processing for collision detection in dynamic environments (STEP2DYNA) under Grant 691154; in part by the Ultra-layered perception with brain inspired information processing for vehicle collision avoidance (ULTRACEPT) under Grant 778062; in part by the China Postdoctoral Science Foundation under Grant 2019M662837; and in part by the National Natural Science Foundation of China under Grant 91648203, Grant U1913601, and Grant U1913205.

**ABSTRACT** The perception of different visual motion cues is crucial for autonomous mobile robots to react to or interact with the dynamic visual world. It is still a great challenge for a micro mobile robot to cope with dynamic environments due to the restricted computational resources and the limited functionalities of its visual systems. In this study, we propose a compound visual neural system to automatically extract and fuse different visual motion cues in real-time using the extremely constrained computation power of micro mobile robots. The proposed visual system contains multiple bio-inspired visual motion perceptive neurons each with a unique role, for example to extract collision visual cues, darker collision cue and directional motion cues. In the embedded system, these multiple visual neurons share a similar presynaptic network to minimise the consumption of computation resources. In the postsynaptic part of the system, visual cues pass results to corresponding action neurons using lateral inhibition mechanism. The translational motion cues, which are identified by comparing pairs of directional cues, are given the highest priority, followed by the darker colliding cues and approaching cues. Systematic experiments with both virtual visual stimuli and real-world scenarios have been carried out to validate the system's functionality and reliability. The proposed methods have demonstrated that (1) with extremely limited computation power, it is still possible for a micro mobile robot to extract multiple visual motion cues robustly in a complex dynamic environment; (2) the cues extracted can be fused with a lateral inhibited postsynaptic network, thus enabling the micro robots to respond effectively with different actions, accordingly to different states, in real-time. The proposed embedded visual system has been modularised and can be easily implemented in other autonomous mobile platforms for real-time applications. The system could also be used by neurophysiologists to test new hypotheses pertaining to biological visual neural systems.

**INDEX TERMS** Bio-inspired, neural network, visual motion perception, micro robot, multiple cues.

## I. INTRODUCTION

Computer vision has underpinned the rapid development of autonomous mobile robots in various applications, such as surveillance, transportation and manipulation [1]–[3]. A distinctive feature of computer vision is that its performance is strictly determined by the scale of available computational

resources [4], [5]. In many remote robotics applications, such as ruin investigation [6], moon rovers [7] and underwater surveillance [8], micro mobile robots often play a unique role because of their size. However, conventional computer vision systems demand a highly capacious computational resource, and may not be readily incorporated into micro mobile robots for autonomous navigation. Novel approaches that function with less computational power are desperately needed for those applications.

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar<sup>1</sup>.

Like robots, many animal species must address quite similar visual challenges in order to navigate (and survive) in a dynamic environment. These challenges are visual motion perceptions which, in animals, are perceived and computed by their extremely efficient visual neural systems. For example, in flying insects, their daily behaviour such as flight stabilisation [9], visual tracking [10], predator detection and avoidance [11], navigation [12] or landing control [13] are controlled mostly by visual cues. Neurophysiological and anatomical studies have shown that these behaviours rely on specific sensory neural pathways. Each of them is highly efficient in extracting specific visual motion patterns in dynamic scenarios with minimum energy costs or structural complexity.

Benefiting from the unique characteristics [14], bio-inspired computational visual perceptive neural systems have been extensively developed and gradually applied in the field of robotics to undertake rapid visual motion recognition tasks. For example, the lateral inhibitory neuron Lobular Giant Movement Detector (LGMD1 and LGMD2) [11], the Small Target Motion Detector (STMD) [15], [16] and the typical lobular plate tangential cells (LPTC) Elementary Motion Detector (EMD) [17], [18] are all highly effective visual neural systems that have been modelled and successfully emulated in small robot platforms.

Although individual neurons can exhibit a very high degree of selectivity for a stimulus produced in response to a specific feature, a single neuron is of limited utility for stimulus encoding [19]. One reason for this is that an individual neuron could exhibit sensitivity to multiple stimuli so that the output response could be mixed and lack precision; another reason is that single neuron's performance fluctuates significantly according to different parameters and scenarios. To eliminate these influences in the vision system of some animals, multiple visual perceptive neurons with common structures and similar neural mechanisms coexist [9], [20], [21]. This mechanism is considered to be a critical prerequisite which enables these species to interact robustly with the dynamic complex real-world. For example, the LGMD1 and LGMD2 are a pair of wide-field detectors in the optic lobe of the locust's which detect looming objects [22]–[25]. The latest research has shown that their neurophysiological behaviours vary significantly despite the fact that their morphological structures have only minor differences. For instance, the LGMD2 selectively responds to darker objects moving on a colliding trajectory against brighter backgrounds whilst the LGMD1 responds to all looming objects [11]. Nonetheless, it is clear that multiple visual cues extracted by different neurons could be the key to the agile reactions and decisions made by animals in relation to real-world dynamic events.

For micro mobile robots, it is critically important that they react to the dynamic environment with rich visual cues, with the highly restricted computing resources. To overcome these challenges, we propose in this paper a “Visual Motion cues Discrimination Neural Network” (ViMDNN) which

functions effectively under extremely constrained computation power, and which enables a micro robot to detect and react autonomously to varied visual motion scenarios in real time. The proposed ViMDNN is constructed in two parts: a presynaptic visual cue detective array and a postsynaptic visual cues fusion neural network, followed by action triggering neuron groups to trigger reaction motion commands. The presynaptic neural array comprises four individual Visual Motion Perception Neurons (VMPN) that detect specific visual motion cues. Each VMPN contains a unified neural structure known as the Extended-LGMD (E-LGMD), which is an optimised neural prototype inspired by the insect LGMD neurons. In the postsynaptic part of the system, visual cues pass results to corresponding action neurons through the feature of lateral inhibition. A Translational Motion Cue Identification System (TMCIS) compares a pair of directional cues to generate translational cues, which are given the highest priority for reactive control, followed by the dark collision cues and approach cues.

The rest of this paper is organised as follows. An overview of related works is given in section II. The proposed models, including the E-LGMD and their postsynaptic connections, are described in section III. Section IV illustrates our experiments on both simulated platforms and real robots. In section V, we further discuss the proposed system and future research.

## II. RELATED WORK

### A. CONVENTIONAL METHODS OF VISUAL MOTION DISCRIMINATION

The majority of conventional methods for visual motion discrimination focus on distinguishing between the physical difference in the foregrounds and backgrounds, including their size, shape and texture details [26]–[28]. So far, preliminary visual motion perception techniques utilise three typical methods: geometric feature detection [29]–[31], background subtraction [32], [33] and optical flow [34], [35].

These methods are powerful in identifying specific objects in complex backgrounds. However, they are highly dependent on prior knowledge of the objects of interest. Even though some learning methods exist to increase the adaptability in unfamiliar and dynamic environments [36]–[38], the growing demand for computational resource is considered as an increasing challenge for real-time processing on a mobile platform [39].

### B. BIO-INSPIRED VISUAL MOTION DETECTORS

Nature provides abundant methods to detect moving objects, which can be classified into looming detectors and lobular plate tangential cells (LPTCs) and Small target motion detectors (STMDs).

Nature provides abundant inspiration for detecting moving objects. The lobular plate tangential cells (LPTCs), LGMD and Directional Selective Neurons (DSN) [40] are generally regarded as typical neurons that respond to wide-field

motion and have been modelled into computational algorithms widely.

The lateral inhibition-based neural models have been applied to robots for many years [41]–[44]. The early models of LGMD1 typically included four layers [41], to which a fifth layer (grouping layer) was eventually added to increase robustness against cluttered backgrounds [43]. Several approaches have been proposed to simulate the selectivity between collision and receding motions [45], [46]. However, the biological basis has not yet been elucidated [11]. The study of LGMD2 by computational modelling commenced only recently and little work has been reported [44], [47]. Fu *et al.* proposed the first LGMD2 model using ON/OFF pathways and Spiking Frequency Adaptation (SFA) to achieve motion direction selectivity [44]. Unmanned Air Vehicles (UAV) are ideal lightweight platforms for LGMD1 and LGMD2 models [48], [49]. In [49], an event camera serves as input sensor to maximise computation efficiency.

There are also other types of bio-inspired neural models which, being simple but sensitive, have been deployed on light-weight platforms. For example, in fruit flies, the well-known LPTC model elementary motion detector (EMD) has been shown to be responsible for detecting translation motion [50]–[52] and has been applied on robots to enable course stabilisation and navigation [17], [53]. The Small Target Motion Detector (STMD) found in dragonflies and hoverflies is specifically sensitive to movements caused by dark objects with very small or limited size [15], and has shown potential in the development of small target automatic detection and tracking systems [16], [54]. Honey bees are widely studied for their flight control behaviour evoked by specific visuo-motor neurons [12], [55], [56]. There is also evidence showing that the praying mantis utilises individual neuro-mechanical visual-motor pathways to control prey-orienting movements [57]–[59]. Most of the afore mentioned neural models exhibit single functionalities, that can barely distinguish between, or recognise multiple visual motion cues.

### C. COORDINATION OF VISUAL MOTION DETECTING NEURONS

Multiple neural models can be integrated to achieve complex tasks. In [53], [60], [61], hybrid models combining both EMD and LGMD systems are proposed. A high-reliable flight control system for UAVs is proposed in [60]. However, these models serve two independent tasks: course stabilisation and collision avoidance. Thus, the system lacks exhaustive recognition of the current visual situation.

Several attempts have been made to combine bio-plausible lateral inhibition models to carry out tasks that require higher level recognition [40], [61]–[63]. As one variant model of LGMD, the DSN is sensitive to translational motion towards a specific direction. DSNs are commonly deployed in an array to identify whole-field translational motion using a competitive mechanism [40], which creates a translational selective neural network (TSNN). Further studies have investigated

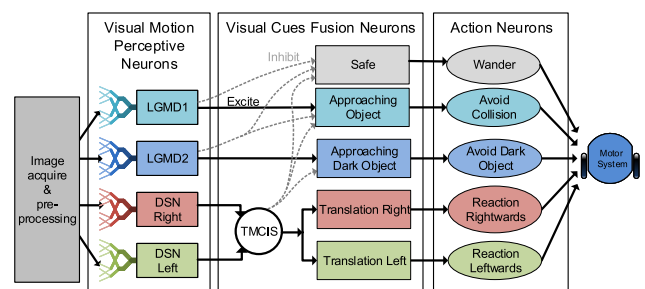
the effectiveness of TSNN combined with LGMD using a computational switching gene [61]. Fu *et al.* proposed a compound visual model that employs both LGMD1 and LGMD2 in competition within one robot to compare their different selectivities [64].

Hu *et al.* proposed a neural structure employing more than ten DSN-like neurons with their corresponding directions arranged in a circle to detect rotation motion detection [62] or even spiral motion [65].

Compared to visual models with single functionalities, these studies demonstrate the potential for recognising dynamic and complex visual motion scenes by integrating multiple bio-inspired visual motion detectors. However, the redundancy in structure could be a significant problem for constrained computing platforms such as micro robot platforms or UAVs.

### III. MODELS AND METHODS

In this section, we present the proposed neural models in detail. The proposed visual system comprises of four sub-systems, including 1) the image capturing and preprocessing; 2) visual motion perceptive neural array; 3) visual motion cues fusion neurons and 4) action neurons, as illustrated in figure 1. Image data is captured and preprocessed first, then transmitted into four individual VMPNs. In the following process, the visual motion cues fusion neurons arbitrate the right visual motion event and select the corresponding pathway to the action neurons. Finally, the results of visual motion perception are indicated by different reactive motor commands autonomously triggered by the micro robot in real-time.



**FIGURE 1.** A schematic of the proposed ViMDNN model. Four VMPNs based on lateral inhibition model (LGMD1, LGMD2, DSNL and DSNR) that serve as visual motion perception neurons are deployed in this system. Their recognised individual visual motion cues are fused and integrated subsequently to generate responding motion commands by the action neurons. The visual cues fusion neurons arbitrate the corresponding output by the defined priority. The dashed arrows represent the inhibition.

As revealed in biological studies, the perception and discrimination of multiple visual motion cues comprises of two stages: 1) the perception of individual visual motion patterns and 2) the fuse and selection of corresponding visual motion cue. In ViMDNN, excitations from all the four VMPNs are gathered and fused. The post-synaptic structure contains two parts: one TMCIS that compares the directional visual motion cues and one array of visual cues fusion

neurons. This bio-plausible structure may reflect an insect’s neural connections that use parallel pathways to process visual information and their mechanisms of natural behaviour generation mechanism [66].

**A. THE VISUAL MOTION PERCEPTIVE NEURONS**

As shown in figure 1, each VMPN model is responsible for a specific visual motion pattern, including the LGMD1/LGMD2 pair that detect approaching motion and DSNs that are sensitive to translational motion. In this work that ground mobile robot is utilised, the left/right translational motion is dominant (caused by ego-motion or objects passing by). In contrast, the up/down translational motion seldom happens. Thus only two DSNs are required. Along with the pair of LGMD to discriminate approaching objects, the final system is composed of four VMPNs, i.e. the LGMD1, LGMD2, left-sensitive DSN (DSNL) and right-sensitive DSN (DSNR).

**1) THE LAYER ARRANGEMENT OF E-LGMD**

The E-LGMD is a layered neural model formed by six layers and several individual neural processing cells and an auxiliary FFI pathway, as illustrated in figure 2. The six layers are the 1) Photoreceptor layer (*P* layer), 2) Excitation layer (*E* layer), 3) Inhibition layer (*I* layer), 4) Pre-Summing layer (pre-*S* layer) 5) Summing layer (*S* layer) and 6) Grouping layer (*G* layer). The luminance change of local individual single eye (pixel) is gathered by each cell in the *P* layer.

The excitation pass into the excitation layer without any delay; however, the inhibition layer accepts delayed signals according to the desired feature. Before the excitations and inhibition meet in the summing layer, an ON/OFF separator is inserted between the excitation/inhibition layers and the summing layer, thus forming a pre-*S* layer. Followed by the summing, a group layer is introduced for signal enhancement. The *K* cell gathers excitations in the *G* layers for generating E-LGMD spikes. The spikes of E-LGMD neuron are generated afterwards.

**2) THE PHOTORECEPTOR LAYER**

In locusts, the first layer of LGMD consists of photoreceptors that represent the excitations from the lamina, which are the luminance change. Similarly, in E-LGMD, the *P* layer is formed by grid-shaped *P*-cells that convey the difference between adjacent frames captured by the camera, with a residue part serving as the visual persistence effect. Respecting to the biological facts, all the VMPNs accept the same visual inputs from a single photoreceptor layer that a single *P* layer acts as their common input. The *P* layer at frame *f* can be defined as a matrix  $P(f)$ :

$$P(x, y, f) = L_0(x, y, f) - L_0(x, y, f - 1) + \sum_{i=1}^{n_p} e^{\mu i} \cdot P(x, y, f - i) \quad (1)$$

where *f* donates the frame index.  $L_0$  is the greyscale image inputs from the camera, which represents the luminance value.  $L_0(f)$  and  $L_0(f - 1)$  are the current and last images. In the visual persistence part,  $n_p$  is the confined steps of persistence. The visual persistence coefficient  $\mu < 0$  determines the decaying speed. The index *i* indicates the last *i*th frames. To balance the performance and complexity, the depth of visual persistence  $n_p = 1$ , and the decaying coefficient  $\mu = -2$ . These parameters are selected empirically with consideration of a balance between the model’s functionalities and the compatibility for embedded environments.

**3) THE EXCITATION AND INHIBITION LAYERS**

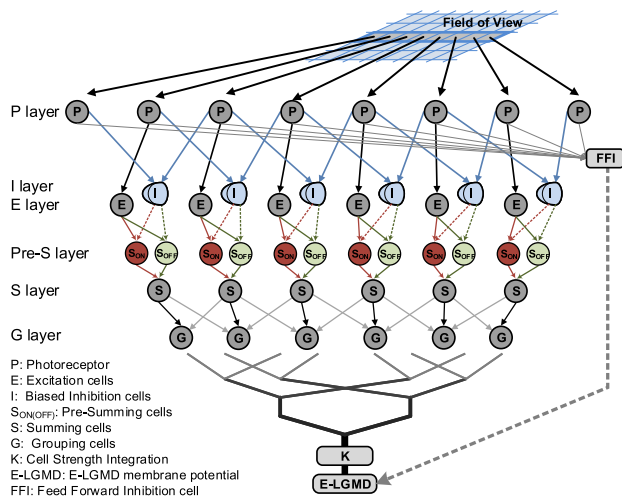
As the core mechanism of E-LGMD modelling, the lateral inhibition is accomplished by utilising two types of layers that exhibit conversed features. The excitation layer (the *E* layer) holds all the current visual motion cues, which are directly retrieved from excitations in the *P* layer.

$$E(x, y, f) = P(x, y, f) \quad (2)$$

The inhibition layer spread inhibitory features with short latency:

$$I(x, y, f) = \omega_I \cdot \sum_{i=-r}^r \sum_{j=-r}^r P(x, y, f - 1) \cdot W_I(i + r, j + r) \quad (3)$$

where the  $\omega_I$  stands for the inhibition coefficient,  $W_I$  is the inhibition kernel pattern. The variation in  $W_I$  represents whether the selectivity is omnidirectional or directional,



**FIGURE 2. A schematic of the proposed VMPN model based on E-LGMD. Notice that the *P* layer is treated as the common part for multiple VMPNs. One column of cells is taken as examples for detailed illustration. The solid arrows indicate exciting pathways, while dashed arrows indicate the inhibition pathways. The *P* cells accept luminance change from the image sensor and pass excitations to the *E* cells directly. The *E* cells obtain value from neighbouring *P* cells. The pathway from the *P* cells to the *I* cells are delayed by one frame. The *I* and *E* cells are then separated by the ON and OFF channels and joined into the pre-*S* cells, then the *S* cells accordingly. The local spatial enhancing mechanism is realised in the *G* layer by grouping a small number of neighbouring *S* cells (indicated by grey arrows). Additionally, for LGMD1, the Feed Forward Inhibition (FFI) gathers excitations from the *P* cells with no delay.**

TABLE 1. Configurations set for the applied VMPNs.

Instance	$n_{cell}$	$\omega_I$	$W_I$	$g_{on}$	$g_{off}$	grouping	$C_\omega$	$t_{de}$	$C_\alpha$	$C_\beta$	$t_{spike}$	$t_{FFI}$
LGMD1	7200 (100x72)	-0.4	$\begin{bmatrix} \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \end{bmatrix}$	1	1	yes	4	30	10	11	100	80
LGMD2	7200 (100x72)	-0.4	$\begin{bmatrix} \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \end{bmatrix}$	0	1	yes	4	30	10	11	100	-
DSNL	5600 (80x70)	-0.53	$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{8} & 0 \\ \frac{1}{4} & 0 & \frac{1}{8} & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	1	1	no	-	20	17	48	-	-
DSNR	5600 (80x70)	-0.53	$\begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{8} & 0 & \frac{3}{8} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & \frac{3}{8} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	1	1	no	-	20	17	48	-	-

which determines the feature of formed VMPN. The  $r$  indicates for the size of applied inhibition pattern  $W_I$ . When pattern size is  $3 \times 3$ ,  $r = 1$ , and for size of  $5 \times 5$ ,  $r = 2$ .

The applied inhibition kernels meet the list in Table 1. Notice that for LGMD1/LGMD2 pair, the kernels are symmetrical, while for DSNs, horizontal biases are applied.

#### 4) THE PRE-SUMMING LAYERS

In the earlier modelling work of LGMD, the lateral inhibition is performed in the  $S$  layer, where the excitation from  $E$  cells and inhibition from  $I$  cells meet. The  $S$  layer in the earlier model is calculated by:

$$S_{RAW}(x, y, f) = E(x, y, f) - I(x, y, f) \quad (4)$$

in which the  $S_{RAW}(x, y, f)$  donates the raw  $S$  cells in earlier simplified models [43], [64]. However, this method should be improved since it cannot separate the signal onset and off-set, which could be the major difference between LGMD1 and LGMD2. In the pre- $S$  layers of the proposed model, the ON/OFF excitations are processed individually. Moreover, concerns about the overflow issue, an anti-overflow limiter is applied to each of the layers:

$$S_{ON}(x, y, f) = [E(x, y, f), E(x, y, f) - I(x, y, f)]^\oplus \quad (5)$$

$$S_{OFF}(x, y, f) = [E(x, y, f), E(x, y, f) - I(x, y, f)]^\ominus \quad (6)$$

in which the  $S_{ON}(x, y, f)$  and  $S_{OFF}(x, y, f)$  stands for the on-channel and off-channel respectively. the definition of  $\oplus$  and  $\ominus$  operations are defined in the appendix. This anti-overflow limiter ensures that the output amplitude of pre- $S$  cells does not exceed the input of  $E$  cells. A simple illustration of how excitation and inhibition affect the pre- $S$  layers are shown in figure 3

#### 5) THE SUMMING LAYER

After joining the excitation and inhibition together into the pair of pre- $S$  layers, the final  $S$  layer is regulated by a switching mechanism.

$$S(x, y, f) = S_{ON}(x, y, f) \cdot g_{on} + S_{OFF}(x, y, f) \cdot g_{off} \quad (7)$$

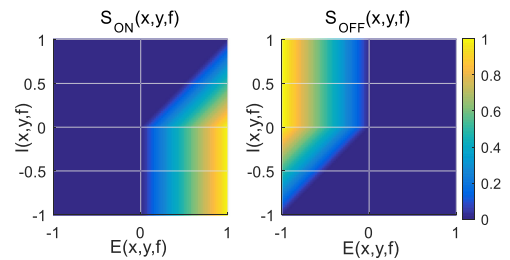


FIGURE 3. An illustration of the process of obtaining  $S_{ON}$  and  $S_{OFF}$  from excitation and inhibition layers, which refers to the eq. 5 and eq.6. The horizontal and vertical axis represent the input value of excitation and inhibition respectively. The output value of pre- $S$  cells' value is indicated by the colour-map.

in which the  $g_{on}$  and  $g_{off}$  are switching parameters that are either 0 or 1. The values for each VMPN are listed in table 1. Notice that the LGMD2 has turned off the ON channel to reject on-sets.

#### 6) THE GROUPING LAYER

The  $G$  layer of the VPMN enhances spatial contrast. The process contains two steps.

In the spatial enhancement process, a passing coefficient, which is determined by local excitation strength, is set for each local cell. The array of passing coefficient  $C_e$  is computed by a 2-D filter:

$$C_e(f) = \sum_{i=-r}^r \sum_{j=-r}^r S(x, y, f) \cdot W_G(i+r, j+r) \quad (8)$$

in which  $W_G$  is the influence from neighbouring cells, which can be described as:

$$W_G = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (9)$$

The spatial enhanced  $G$  layer is then obtained by:

$$G(x, y, f) = S(x, y, f) \cdot C_e(x, y, f) \cdot \omega(f)^{-1} \quad (10)$$

where  $\omega$  is a scale that measures the passing coefficient intensity of the whole frame:

$$\omega(f) = 0.01 + \max \left| Ce(f) \cdot C_\omega^{-1} \right| \quad (11)$$

where  $C_\omega$  is a certain value.

In the second step, small signals are blocked by a certain threshold:

$$\tilde{G}(x, y, f) = \begin{cases} G(x, y, f) & \text{if } G(x, y, f) \geq t_{de} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where the  $t_{de}$  is the constant threshold that filters out small signals. For modelling DSN, only the second process (eq.12) is involved since grouping will increase undesired non-linearity to later process in TMCIS.

#### 7) THE SPATIAL INTEGRATION NEURON

The spatial integration neuron (K cell) in this VMPN model represents the average excitation levels in the previous layer. It first gathers neural excitations from all cells in the previous layer:

$$k(f) = \sum_x \sum_y \left| \tilde{G}(x, y, f) \right| \quad (13)$$

Following by that, the value  $k_f$  is transformed by a normaliser:

$$\kappa(f) = \frac{1}{2} \cdot \tanh \left( \frac{\sqrt{k(f)} - n_{\text{cell}} \cdot c_\alpha}{n_{\text{cell}} \cdot c_\beta} \right) + \frac{1}{2} \quad (14)$$

where the  $c_\alpha$  and  $c_\beta$  are constants that can shape the normalising function with varied gain settings for small and big signals. The  $n_{\text{cell}}$  is the number of element cells in the  $G$  layer.

The structures followed by this K cell is dependent on the actual model. In this work, the output of four utilised VMPNs are represented by  $\kappa_1(f)$ ,  $\kappa_2(f)$ ,  $\kappa_L(f)$  and  $\kappa_R(f)$  respectively. For the LGMD1/LGMD2 model,  $\kappa_1(f)$  and  $\kappa_2(f)$  cells are followed by the spiking mechanism that represent a prominent approaching event. For DSN modelling, the comparison of  $\kappa_L(f)$  and  $\kappa_R(f)$  will be described later for the TMCIS modelling.

#### 8) THE SPIKING MECHANISM FOR LGMD1/LGMD2

In E-LGMD process, if the spatial integration neuron exceeds a determined threshold, a spike is produced:

$$s_{\text{spike}}(f) = \begin{cases} 1 & \text{if } \kappa(f) \geq t_{\text{spike}} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

An impending collision is confirmed after several successive spikes. Longer decision time lead to solid and reliable outputs, but it is harder to respond to sudden and fast collision events. In our tests, the decision time is set to 3-6 frames (100-200 ms) for best performance.

#### 9) THE FFI PATHWAY OF LGMD1

In LGMD1 circuitry, an auxiliary feed-forward inhibition pathway is responsible for detecting large and sudden whole-field visual motion, which could be introduced by self-rotation or other unknown significant motion. The FFI neuron and lateral inhibition work together to prevent false spiking alarms in these situations. When the VMPN is configured into other neuron types, the FFI mechanism is bypassed.

The FFI cell is proportional to the average excitation level of  $E$  layer with one frame delay:

$$k_{\text{FFI}}(f) = \sum_x \sum_y |E(x, y, f - 1)| n_{\text{cell}}^{-1} \quad (16)$$

A constant threshold is set to enable the FFI mechanism. This threshold should be high enough that it will not interfere with normal reactive control.

$$s_{\text{spikeFFI}}(f) = \begin{cases} 1 & \text{if } k_{\text{FFI}}(f) \geq t_{\text{FFI}} \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

#### 10) THE PARAMETERS SETTING

By setting the unified E-LGMD models with different parameters, they can be initialised into varied VMPNs with unique functionalities. This facilitates the implementation of multiple models with little extra ROM occupation. The parameters' configuration of applied VMPNs are described in table 1. Currently, there is no learning or adaptive mechanism applied in E-LGMD thus most of the parameters are empirically fixed considering the functionalities of designed models. However, there are several methods developed to optimise the parameters [67]. As mention before, some parameters are selected for better compatibility with embedded environments. Notice that the  $W_I$  are fractions with denominators of 4 or 8. Meanwhile, some parameters are determined by the physical properties of the model such as the value of  $n_{\text{cell}}$  and the gating parameters.

#### B. THE POSTSYNAPTIC NEURONS

As a higher-level architecture, the ViMDNN coordinates multiple VMPNs according to their functions and generates the right behavioural response to each type of visual challenge. Therefore the robot with the ViMDNN can cope with different visual stimuli reliably. If only with single function model, for example, only with LGMD, the robot will not be able to respond to translating objects correctly; only with DSNs, it can hardly cope with collision events. The ViMDNN provides the right structure for a robot to initiate the right response to moving visual cues in the real world.

However, different VMPNs cannot simply be integrated together. The LGMD produce intensive spikes when an object approaches. This spiking mechanism contributes to rapid and solid results but with strong non-linearity. On the contrary, the DSNs are tuned to be sensitive to certain motion directions, although they also exhibit minor excitations to approaching objects or even translating objects towards other directions. Therefore, the DSNs must be merged and

transformed into spiking outputs before they meet the LGMD neurons. The TMCIS is deployed to achieve this goal.

1) THE TRANSLATIONAL MOTION CUE IDENTIFICATION SYSTEM

For flying insects such as fruitflies, two steering strategies are applied, which are course stabilisation and obstacle avoidance. Required visual information could be obtained by comparing between left and right visual motion trends through the visual motion detector EMD [68]. It is demonstrated that when triggering saccade reactions, the direction of saccades (left or right) is opposite to the side who experience larger visual motion. This approach has also inspired several controlling models in robotics [69], [70] with EMD neurons.

Similarly, a TMCIS based on E-LGMD neurons can provide directional information for steering control of robots, compromising results from the divergence of a pair of DSNs. The output of TMCIS represents the dominating motion direction. The structure of TMCIS and the comparison to fly’s Spatial-temporal Integration if Motion (STIM) model is illustrated in figure 4.

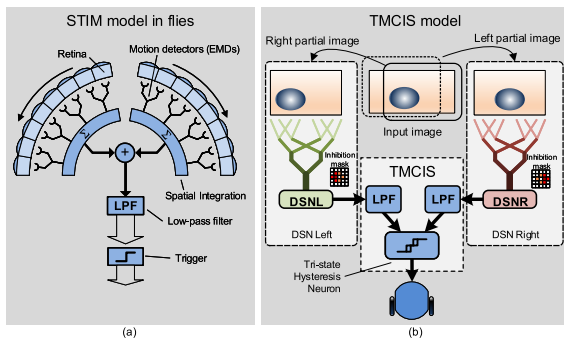


FIGURE 4. (a) The spatio-temporal integration of motion (STIM) model inspired by fruitflies (figure adapted from [71]). (b) A schematic of the proposed TMCIS model. Notice the difference of inhibition masks in DSNL and DSNR.

The basic idea of obtaining the directional motion is to compute the divergence of the DSN cells directly:

$$d_{RAW}(f) = \kappa_L(f) - \kappa_R(f) \quad (18)$$

in which the output value  $d_{RAW}(f)$  contain the information of both the direction and strength of translational motion. However, due to the noise introduced by calculating DSN, this approach is not practical and should be improved by further noise reduction and gating techniques.

Here in TMCIS, it evaluates the divergence between the pair of DSNs, namely  $d_{TMCIS}(f)$ . First, the output of both DSNs are filtered by a single-pole recursive filter (also can be regarded as a leaky integrator) to reduce noise:

$$\tilde{\kappa}_L(f) = \tilde{\kappa}_L(f - 1) \cdot \eta_L + \kappa_L(f) \cdot (1 - \eta_L) \quad (19)$$

$$\tilde{\kappa}_R(f) = \tilde{\kappa}_R(f - 1) \cdot \eta_R + \kappa_R(f) \cdot (1 - \eta_R) \quad (20)$$

where  $\eta_L$  and  $\eta_R$  are the decaying factors within range (0,1). A smaller value contributes to a lower filtering strength.

A greater value contributes to smoother output but with longer latency. This is a practical approach of low-pass filters for micro-controllers. In TMCIS, both decaying factors are empirically set to 1/7.

Then the divergence of two DSNs  $d_{TMCIS}(f)$  is calculated by:

$$d_{TMCIS}(f) = \frac{(\tilde{\kappa}_L(f) - \tilde{\kappa}_R(f))^2}{\sqrt{\tilde{\kappa}_L(f)^2 + \tilde{\kappa}_R(f)^2}} \cdot \text{sign}(\tilde{\kappa}_L(f) - \tilde{\kappa}_R(f)) \quad (21)$$

where the strength of the divergence ( $|d_{TMCIS}(f)|$ ) shows the strength of dominate direction of motion. This operation slightly compresses small signals to increase the robustness of direction detection, because a small signal means the  $\kappa_L$  and  $\kappa_R$  grows simultaneously, which is not the case of translational motion. A simple illustration of this mechanism is shown in figure 5.

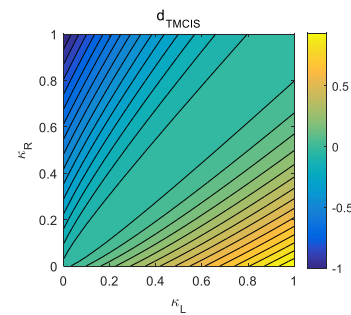


FIGURE 5. The illustration of obtaining  $d_{TMCIS}$  from the outputs of DSNs. Notice that the output is suppressed (flattened) along the line:  $\kappa_L - \kappa_R = 0$ .

a: THE COMPARING CELL

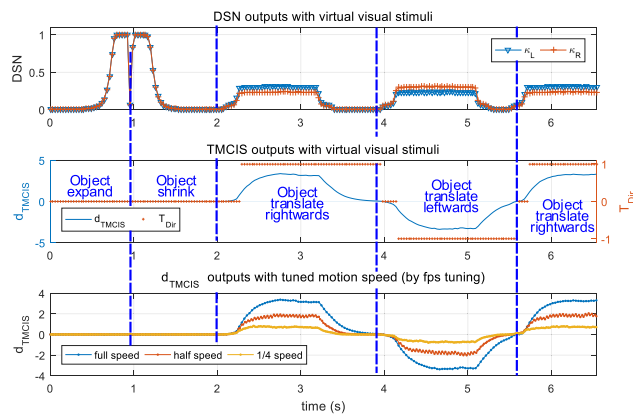
A solid decision of translational motion is made when  $d_{TMCIS}(f)$  has steady output. A tri-state hysteresis neuron is applied here to reduce sudden and frequent fluctuations. The tri-state hysteresis neuron has two constant on-set thresholds: one on each side and an off-set threshold zero, which contribute to three outputs: left, right and centre. They are represented by  $-1$ ,  $1$  and  $0$  respectively. The decision of a certain direction is selected when the input strength (absolute value) is greater than the on-set threshold, however, it will return to the centre state once the input strength drops below the off-set threshold and cross to the other side. When the value is between two thresholds, the decision remains unchanged:

$$T_{Dir}(f) = \begin{cases} 1(right) & \text{if } d_{TMCIS}(f) > t_{TR} \\ -1(left) & \text{else if } d_{TMCIS}(f) < t_{TL} \\ 0(centre) & \text{else if } d_{TMCIS}(f) \cdot T_{Dir}(f - 1) \leq 0 \\ T_{Dir}(f - 1) & \text{otherwise} \end{cases} \quad (22)$$

in which  $T_{Dir}(f)$  is the decision at frame  $f$ . The  $t_{TR}$  and  $t_{TL}$  are two thresholds for the left side and right on-set ( $t_{TR} > 0$ ,  $t_{TL} < 0$ ). In this work, the values are set to  $t_{TR} = 4$  and  $t_{TL} = -4$ .

### b: THE CROSS PARTIAL IMAGE CONFIGURATION

Consider the situation that a right motion happens in the rightmost part of the view. It is usually not hazardous since the object has almost moved outside the field of view. Similarly, the leftmost part of the view can be skipped when judging a hazardous left-moving object. Thus, the input of DSNs can be tailored to increase selectivity and reduce processing time. The DSNL takes only the right part of the input image into the process, and the DSNR calculates only left part. In TMCIS model, the overlap of 60% provides good separation, as shown in figure 4. A simple test of TMCIS with video stimuli is shown in figure 6.



**FIGURE 6.** Test of TMCIS with a virtual visual stimulus cast on the screen. The virtual stimulus' motion trajectory is approaching (expand)  $\rightarrow$  receding (shrink)  $\rightarrow$  moving right  $\rightarrow$  moving left  $\rightarrow$  moving right. The upper figure shows the DSN outputs. The middle figure shows the  $d_{TMCIS}$  and  $T_{dir}$ . The lower figure shows the  $d_{TMCIS}$  with different motion speeds.

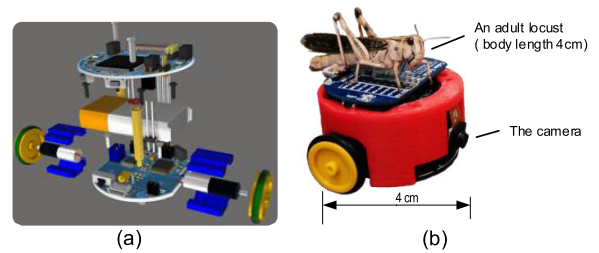
### 2) THE VISUAL CUES FUSION NEURON

As illustrated in figure 1, the proposed visual cues fusion neurons collect visual motion cues from VMPNs and the TMCIS, then generate a single output directly to the action neurons. The extracted visual motion cues could be either of: 1) safe, 2) approaching motion, 3) approaching motion caused by dark objects, 4) translational motion to the left and 5) translational motion to the right.

The visual cues fusion neurons use a prioritised preemption mechanism to organise the orders of input neurons. In this work, five cells representing possible identified visual motion situation are assigned with defined priority to connecting presynaptic neurons, including a default cell. A neuron can inhibit other neural excitations whose priority is lower to ensure only one output is active at any time. The LGMD1 and LGMD2 are responsible for detecting approaching object motion, which has a higher priority than the safe situation. Further identification of the dark approaching object can be given by LGMD2. However, the decision of an approaching motion will be inhibited by a translational motion (either left or right) given by the TMCIS.

### C. THE SYSTEM IMPLEMENTATION AND OPTIMISATION

According to the structure of the proposed neural model, only low-level calculations are involved in the whole process



**FIGURE 7.** The utilised micro-robot platform Colias-IV. (a) An explosion view of the mechanical structures, including the boards, battery and motors. (b) An assembled robot compared the size to a genuine adult locust. They are nearly equal in length.

such as excitation transferring and neighbouring operations, which make it a desired model for embedded platforms with constrained computational resources. However, it is still not an easy task for adapting this neural structure successfully.

### 1) THE ROBOT PLATFORM

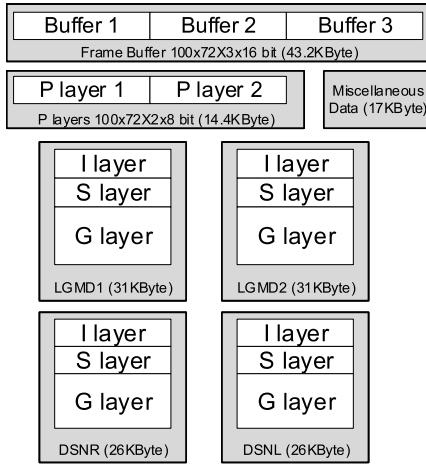
A micro ground robot Colias-IV is selected, as shown in figure 7, to be the application platform to demonstrate the high computing efficiency and low-hardware demanding. The Colias-IV which has constrained computation resources is a low-cost vision-based micro ground robot developed for swarm robotic applications and bio-robotic research [72]. The Colias-IV employs a circular body with 4cm in diameter that has two layers of circuit boards. The bottom part serves as an only primary robotic platform such as power supply, motion control and primary sensing. Driven by a pair of differentially driven wheels and a third pin-stand in front, the Colias-IV can run up to 35cm/s at full speed. The upper part, which called the Colias-Sensing-Unit (CSU), is designed for on-board high-level sensing such as vision. The CSU contains a tiny CMOS camera as a visual sensor that captures RGB image sequences at 30 frames per second (fps). The field of view angle (FOV) is about 70 degrees horizontally. The CSU utilises an STM32F427 microcontroller as the main processor, which is running at 180MHz. It contains 256KBytes internal RAM space and 2MBytes ROM space.

### 2) THE COMPUTATIONAL COMPLEXITY ANALYSIS

For embedded programming, both spatial complexity and temporal complexity are critical issues and should be optimised. In addition to the basic principle of optimising ELGMD-based model described in [42], the  $P$  layer is regarded as a common part for all the VMPNs for further memory saving. Thus the  $P$  layer can be stored separately and calculated only once for each frame. Regarding the calculation procedure, two consecutive  $P$  layers should be accessible for implementing delayed inhibition, and for differential image calculation and temporary storage, three input buffers are allocated. Due to the maximum available RAM space, the utilised scale of the input image is 100x72. The breakdown of total occupied memory space is shown in figure 8.

The time used for each process also depends on the scale of active neurons. For each E-LGMD, even though five layers





**FIGURE 8.** The illustration of RAM space allocation for images and E-LGMD models. The total occupied RAM is around 190KBytes. Notice that two P layers are independent and are re-used by all E-LGMD models. In each E-LGMD model, the size of G layer (16-bit signed number) is almost twice than the I layer and S layer (8-bit unsigned).

exist, calculations for some layers can be omitted or merged in a single loop to save time. As briefly illustrated in algorithm 25 and figure 9, the calculations of I layer and S layers are combined into one loop, the calculation of Ce map and the G layer are also combined. Meanwhile, the boundary of convolution is omitted that the time cost for convolution operation can be further reduced by skipping the edge cells. Thus, the computation process of the full process mainly consists of three loop phases: 1) the P layer, 2) I layer and S layer of each VMPN, and 3) the loop for G layer and spatial integration for each VMPN. When the input image has the size of  $H_{in} \cdot W_{in}$  pixels and convolution kernel size of  $k$ , the computational complexity of each phase can be estimated as  $N_P = O(H_{in} \cdot W_{in})$ ,  $N_{IS} = O(k^2 \cdot (H_{in} - 2) \cdot (W_{in} - 2) + H_{in} \cdot W_{in})$  and  $N_G = O(k_c^2 \cdot (H_{in} - 4)(W_{in} - 4))$ , where  $N_P$ ,  $N_{IS}$  and  $N_G$  represent time for the three phases.

$$N_{VMPNs} = N_P + 2N_{LGMD} + 2N_{DSN} \quad (23)$$

where  $i$  represents the index of utilised VMPNs. A test of time consumption in different image scales is shown in figure 10a. The timing profile of utilised model that has a scale of  $100 \times 72$  is shown in figure 10b.

As a result, these further optimisations, the total time used for processing all the modules including four E-LGMD structures has reached  $23 \pm 2$  ms, which is within the duration of one single frame ( typically 33ms). This indicates that the proposed compound visual motion detection system can be performed in real-time on the Colias-IV platform.

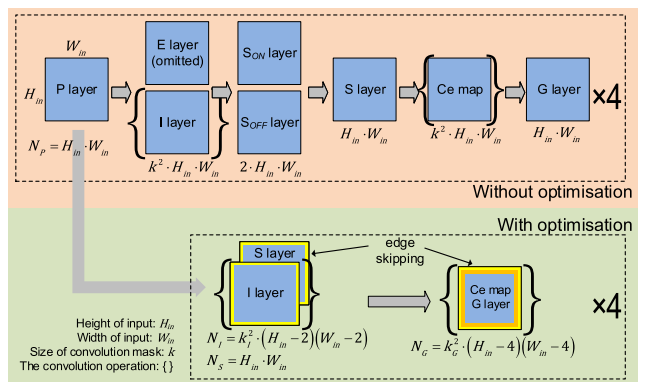
The power consumption of the model is mainly depended on the MCU's clock frequency due to the architecture of the embedded processor. In its typical condition (180MHz), the power consumption of MCU is around 76mA with the capability of processing 225 Dhrystone Million Instructions Per Second (DMIPS). The power consumption breakdown is briefly listed in table 2.

**Algorithm 1** The Process of ViMDNN

```

1 while has valid image do
2   // Calculate VMPNs
3   Process the P layer;
4   foreach E-LGMD do
5     // First cell-wise loop
6     if not to omit E layer then
7       deal with E layer;
8     Process I layer;
9     if is LGMD1 then
10      Calculate FFI;
11    if is not LGMD2 then
12      Calculate SON;
13    Calculate SOFF;
14    Calculate S layer;
15  foreach cell do
16    // Second cell-wise loop
17    if is LGMD1 or LGMD2 then
18      Process Ce value map;
19      Calculate G layer;
20      Process G layer thresholding;
21    else
22      Process simple G layer thresholding;
23    Process Spatial Integration ;
24  Process Spiking Mechanism ;
25  Calculate TMCIS;
26  Calculate VCFM;
27  Motion Control;

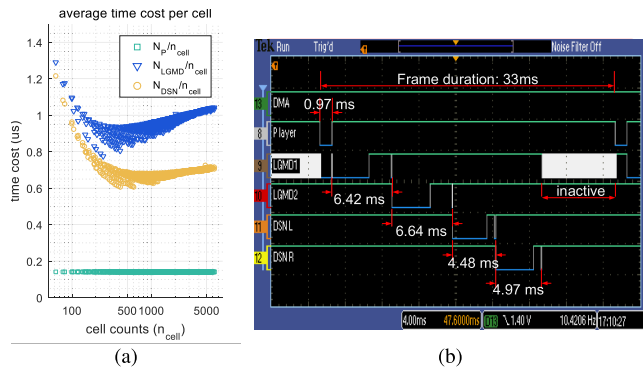
```



**FIGURE 9.** The method to optimise for computation time. Comparing to the upper part which is a straightforward implement without optimisation, the optimised algorithm 1) re-use P layer for all VMPNs; 2) The merging of layers into two main loops; 3) the edges of convolution operation are skipped. Notice that in G layer, two boundary rings are skipped.

**IV. EXPERIMENTS AND RESULTS**

The performance of proposed ViMDNN is evaluated in systematic experiments including basic functional tests with virtual visual stimuli, real-world challenges and simple autonomous control scenarios.



**FIGURE 10.** Timing profile of real-time processing on the embedded processor. (a) The measured time cost in different phases per cell when the size of input image changes, showing the measured results of LGMD1 and DSNL as examples. The x-axis is in log scale. Notice that in the most efficient situation, the input size should be 400-600 pixels. (b) The timing profile observed and logged by a multi-channel oscilloscope (Sampling Rate 1GHz) using 100 × 72 cells. The time measuring labels are added manually. In the duration of a full frame (33 ms), four VMPNs are processed successively after the P layer calculation, and the whole execution is completed within 24ms. The change of calculation phases is indicated by the signal flips.

**TABLE 2.** The power consumptions of major components.

Module Name & Description	typical	max	unit
MCU standby *	2.5		
MCU core (150MHz) **	46	52	
MCU peripherals (150MHz)	13	15	
MCU core (180MHz)	57	65	
MCU peripherals (180MHz)	19	25	
Camera standby	0.8		<i>mA</i>
Camera active	12	18	
CBU (without motor)	20	75	
DC Motor x2	60 ***	300	

\* This set the MCU into freeze mode.

\*\* The main clock can be configured at start-up by setting the inner registers.

\*\*\* The robot runs at 8cm/s under this condition.

**A. RESPONSES OF VMPNs TO VIRTUAL VISUAL STIMULI**

The proposed VMPN models are firstly tested by virtual visual stimuli to demonstrate their fundamental features. The computer-generated visual stimuli are categorised into two groups: 1) an approaching object that has an expanding edge and 2) a translating object that moves from left to right with constant speed, then reverse. Each group of visual stimuli is set with different object-background contrasts  $C_{obj}$ , which is defined by:

$$C_{obj} = B_{obj} - B_{back} \tag{24}$$

where  $B_{obj}$  is the brightness value of moving object,  $B_{back}$  is the brightness of background. The brightness values are within the range [0,1]. No additional noise is added to the image stimuli. Experiment stimuli, explanation and sample data logged from the robot are shown in figure 11.

**1) SIMULATED APPROACHING OBJECTS**

Firstly, we would like to compare the selectivity between LGMD1 and LGMD2 when an object is on the approaching trajectory, as shown in figure 11a. When challenged by a dark approaching object (figure 11e up-left), both LGMD1 and LGMD2 generate increasing membrane potential rapidly as the object expands in the retina. When the object is bright while the background is dark (figure 11e up-right), the LGMD2 neuron is inhibited during the whole approaching process as expected. The reactions of DSN towards virtual approaching visual stimuli are also tested, as shown in figure 11b. Since the object on an approaching trajectory donates brightness change on both sides evenly, it is clear that both DSNs generate spikes evenly and show no direction preference on any object/background contrast group.

**2) SIMULATED TRANSLATING OBJECTS**

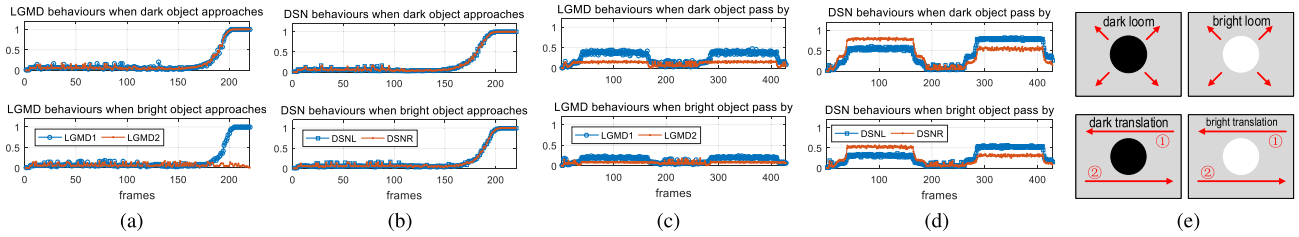
When the simulated visual stimulus is moving at translating trajectory (figure 11e down), the LGMD1 and LGMD2 neurons show only constant and weak outputs during the whole process for either dark or bright stimulus (figure 11c). In this situation, the peak values for both motion directions (left and right) is almost the same, indicating that the pair of LGMD neurons is not sensitive to motion directions.

Meanwhile, the outputs of both DSNs differ along with the motion direction. For the motion from left to right regardless of the object/background contrast, the DSNL is weaker than DSNR, and vice versa. The divergence between these two DSNs, which is close to the TMCIS output described in the above section, have almost the same strength for motion in two opposite directions, The results demonstrated the functions of biased inhibition mechanism set in the *I* layer.

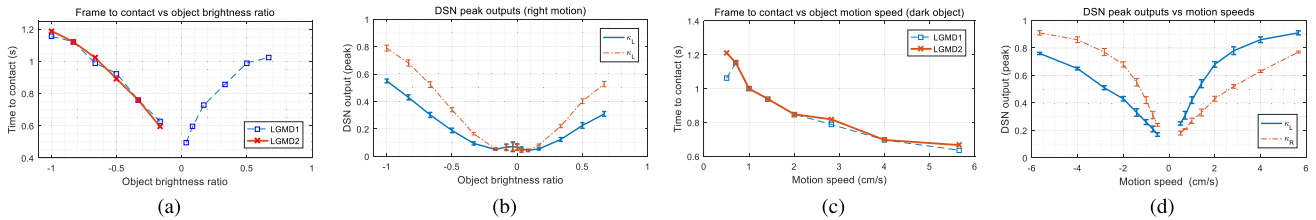
The results of tests on simulated visual stimuli with different object/background contrast set are compared and concluded in figure 12. The results shown in figure 12a show that LGMD2 only detects an approaching object when it is darker than the background. Moreover, the sensitivity of proposed LGMD1 and LGMD2 structures decrease when the object becomes more obscure against the background, resulting in a shorter alarm time for responding reactions.

The peak output of DSNs in the object translating scene has shown stable and distinctive selectivity within a single motion scene, as illustrated in figure 12b. It shows that the sensitivity of each DSN is related to the object/background contrast. The clearer the object stands out from the background, the higher values generates. Results of additional tests on translating objects from the opposite side have shown little difference in the peak value, except the swapping of places. This result reveals that the peak value of DSN is related to the contrast with the given direction, but the polarity of motion does not affect the performance of DSN, which is different from the LGMD2 model.

Figure 12c has drawn a relationship between the time to collision (TTC) and the motion speed. When the motion gets faster, the TTC decreases, which result in a shorter



**FIGURE 11.** The experiments of VMPNs challenged by computer-generated visual stimuli. (a) & (b) The output of both LGMD1, LGMD2 DSNL and DSNR when the virtual stimulus is on an approaching trajectory. (c) & (d) The output of all VMPNs when the virtual stimulus is translating from one right to left, then left to right. (e) The explanation of applied virtual visual stimuli. The colour combination of both background and foreground are tested in groups.



**FIGURE 12.** Analysis of experiments on virtual stimuli. (a) The TTC of both LGMD1 and LGMD2, when challenged with approaching objects under different object/background contrast settings. Notice that the LGMD2 only respond to dark objects in front of a bright background; (b) The peak value of DSNs when challenged with translating objects from left to right. Showing the standard errors of logged data during the progress; (c) The TTC of LGMD1 and LGMD2 when the moving speed of approaching object changes. (d) The peak value of DSNs when challenged with translating objects with different motion speeds, where  $C_{obj} = -1$ . Showing the standard errors with errorbar.

responding time for an agent to take actions to avoid an impending collision.

The relationship between DSN outputs and the translating motion speed is also tested, as shown in figure 12d. The results show that both the DSNs output increase along with the moving speed, but selectively gradually saturates when the speed of the object increases.

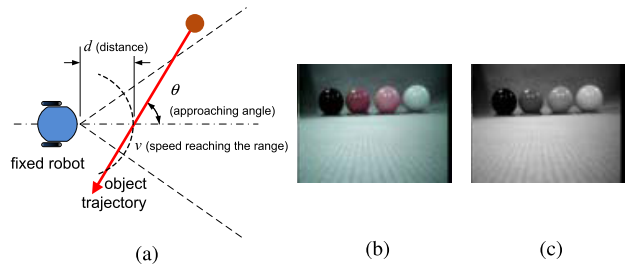
**B. EXPERIMENTS WITH REAL-WORLD STIMULI**

Comparing to computer-generated virtual stimuli, the visual scene in real-world is noisy, caused by unwanted flashings or shadows cast by poor illumination. Moreover, the object’s motion speed cannot be strictly maintained as video do. Therefore, the performance of proposed neural models should be tested in the “real” dynamic scenes.

**1) CHALLENGE WITH ROLLING BALLS**

To test the postsynaptic visual cues fusion neurons and demonstrate the ability to distinguish different motion patterns successfully, a rolling ball is used as the visual stimulus in the following experiments.

The experimental configuration used in this test is shown in figure 13. Several billiard balls with four different pure colours are used as visual stimuli, which the variety in colours contribute to different brightness levels in the luminance channel ranging from totally black to very light grey. The object’s trajectory is defined by the approaching angle against the robot’s median plane  $\theta$  and a determined safe range  $d$ . The object’s speed  $v$  is approximately 20 cm/s when crossing the robot’s median plane. Sets of experiments are conducted



**FIGURE 13.** The set-up for object looming experiments. (a) The illustration of defining the looming object’s trajectory. (b) The objects used in the experiments, with four different colours. The image is captured by the robot itself, stored with the original YUV format and displayed without alteration. (c) The brightness channel of (b) that motion detection models actually processes, which indicates the visual stimuli and background’s real appearance of brightness levels.

in nine groups of motion trajectories, which are listed in table 3. For each group of motion trajectories, experiments are repeated for 20 times. During the object moving process, the recognised visual motion cues from the compound model are logged and compared in figure 14. The result of motion pattern detection is shown in figure 15 and figure 16.

The neural outputs during the tests are measured, and the response details in some typical scenarios are illustrated in figure 14. In the results, the performances of designed VMPN models are consistent with our expectations. Both LGMD1 and LGMD2 can recognise dark approaching objects and produce spikes accordingly, but the LGMD2 produces slight excitation when a bright object is observed. When the object is near-miss but not on a head-on collision trajectory, both LGMD1/LGMD2’s response cannot reach

TABLE 3. Experiment configurations of looming balls.

scene	$d$ (cm)	$v$ (cm/s)	$\theta$ ( $^\circ$ )	description
1	-	20	0	head on collision
2	50	20	$\pm 15$	near miss
3	10	20	45	cutting through (right)
4	10	20	90	translating (right)
5	10	20	135	overtaking (right)
6	10	20	-45	cutting through (left)
7	10	20	-90	translating (left)
8	10	20	-135	overtaking (left)
9	35	20	90	translating far away (right)

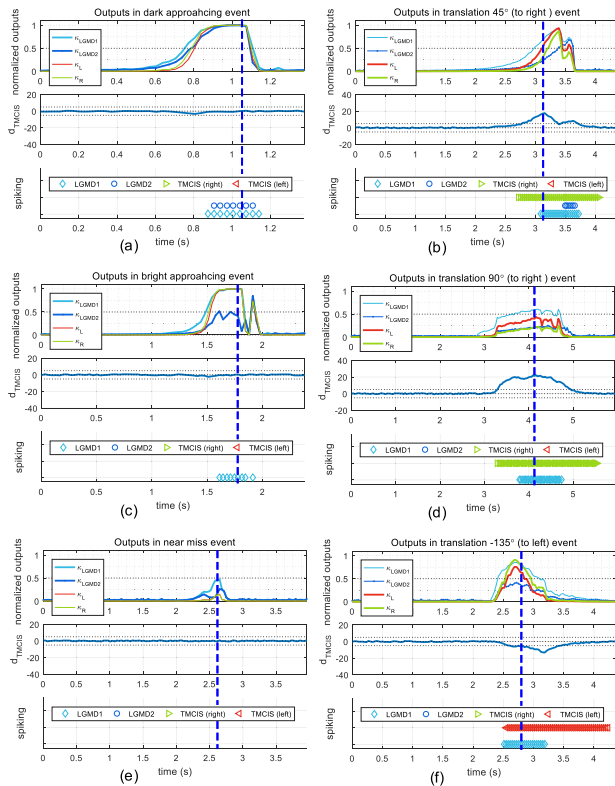


FIGURE 14. Model variables and outputs during a typical experiment with real-world stimuli, showing the VMPNs' outputs (upper plot), TMCIS output  $d_{TMCIS}$  (middle plot) and spiking outputs (lower plot). In the output plots, horizontal dashed lines indicate the threshold for shown models. Vertical dashed lines indicate when a collision happens, or the stimulus cross the median plane. The decision of recognised motion is made by the uppermost spiking neuron shown in the spiking plot.

the threshold to produce any spikes. The TMCIS formed by two DSNs can successfully recognise translating objects in certain scenarios. In object approaching scenarios, although both DSNs are triggered vigorously, their divergence is too small to contribute a sound output in TMCIS. The decision making of visual cues fusion neurons is also illustrated in the figure. The uppermost spike is the corresponding fused neural output that represents the current visual situation.

The recognised visual motion results are briefly counted in figure 15, which shows the decisions with a head-on colliding object with varying brightness levels. For white

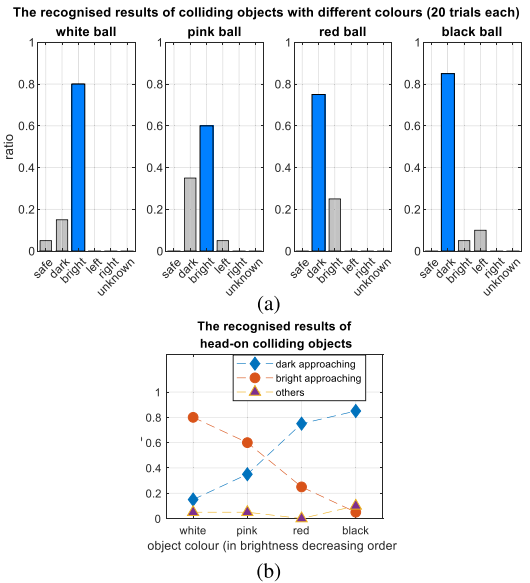


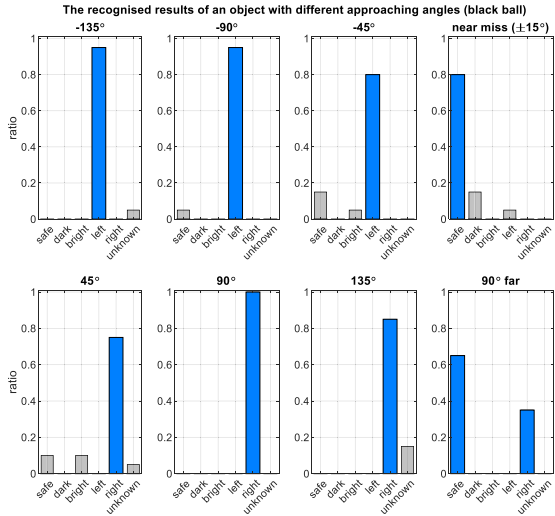
FIGURE 15. The performance of motion pattern recognition evoked by a head-on colliding object with varied colours. (a) The counted results of recognised motion patterns of the head-on colliding object. The CORRECT decision is marked with a blue bar. (For white and pink objects, the bright approaching results are considered as "correct". For red and black objects, dark approaching are considered as "correct".) The maximum chance of correctly recognise a motion pattern is 1 (100%). (b) The ratio of detected bright approaching and dark approaching results vs. different ball colours.

and pink visual stimuli that are brighter than the background, the ViMDNN model showed more bright approaching results than rest of the decisions, however, for red and black visual stimuli, the dark approaching result is preferred. The figure 15b showed that the object's brightness affects the probability of what category the object would be classified. The total error rate (other decisions other than two types of collision) is less than 10%.

In figure 16, the results of challenges with a single black ball but on different translating paths are shown. For objects moving not far away from the robot, most of the translating motion can be recognised and separated correctly. However, for the motion happening far away from the robot, the scene has a higher chance to be treated as safe rather than a translational motion. Notice that in some cases, the ball suddenly appears and passed by, an unknown scene might be prompted due to the excitation of FFI cell in LGMD1 brought by the fast brightness fluctuations.

## 2) EXPERIMENT WITH ROBOT BEHAVIOUR CONTROL

In the experiments above, it has been demonstrated that each of the VMPNs can reliably detect specific motion patterns, and the visual cues fusion neurons, including the TMCIS, can integrate and fuse the visual cues and make a final decision. Now the motor commands are associated with each recognised visual motion situation to test whether the integration of VMPNs in a miniature mobile robot can trigger responding actions robustly.



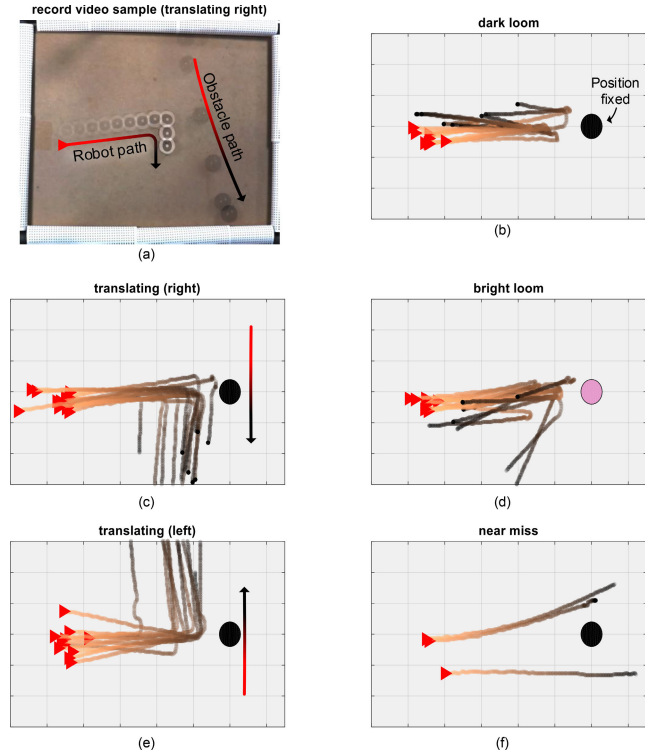
**FIGURE 16.** The ratio of recognised results stimulated by a black ball with different motion trajectories (approaching angles). The definition of motion trajectories is listed in table 3. For objects translating 90° far away (last figure), both translating and safe decisions are considered as correct.

An experiment arena is built for the Colias-IV robot to run freely inside. The arena is about 70x60cm<sup>2</sup> surrounded by low barriers which leave the robot’s view unblocked to the open laboratory environment. An obstacle (a billiard ball described in previous experiments) is used as the visual stimulus. The path of the obstacle can be controlled when repeated.

The robot is set free from a fixed position inside the arena. Before it detects an obstacle in front or translating in front, it remains a certain course. Four kinds of visual events are employed to challenge the robot: 1) fixed position dark loom; 2) fixed position bright loom, 3) object translating from left to right and 4) object translating from right to left. Testing scenarios are repeated 50 times each. The obstacle used in the experiment is a black ball as dark loom and a white ball as the bright loom. The motion command rules responding to the visual obstacle is kept simple but clear enough to be recognised, as described in table 4.

The robot’s path with correct responding to the testing event in this experiment are illustrated in figure 17, and the data conclusion of the experiment is illustrated in figure 18. A top-down camera tracks the robot’s path by the Whycon [73] tracking marker carried on top of the robot.

From the observed results, the following analysis can be given, that 1) the proposed ViMDNN model can detect most of the dynamic visual motion events and corresponding trigger reactive motor commands. 2) The average success rate (as indicated in figure 18 ) of detecting a visual stimulus correctly with our proposed systems is slightly more than 85%. 3) The Distance to Collision (DTC) is nearly constant, which demonstrates the robustness of the ViMDNN model. However, evidence from the previous study proves that DTC would increase as the moving speed gets faster [42], while the TTC decreases that allowing less time to respond to a certain event, according to the results depicted in figure 12c.



**FIGURE 17.** Illustration of some typical robot behaviours responding to each kind of recognised visual stimuli. The experiments are recorded by a camera above the arena. (a) The recorded video during a right translating event to show the experiments’ set-ups, showing the paths of the robot and the obstacle during the period; (b-f) The typical response to certain visual stimuli showing the robot path and the position (trajectory) of the obstacle. The start point of the robot in each trace is marked by a red triangle. The definition of assigned motion command is listed in table 4. Particularly in (f), shows the occasionally occurred near-miss trajectories when conducting experiments.

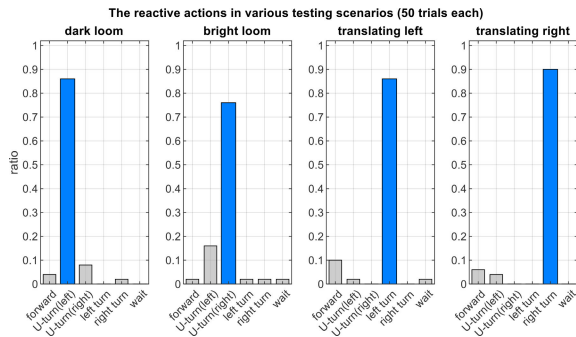
It should be noted that compared to the study of a single visual model on a micro robot such as LGMD2, it is difficult to present different types of visual stimuli recognisable and countable rather than looming only in a long-term arena run setting. Within an arena, the proposed method may be able to demonstrate the robustness in responding to looming, it can hardly show the discrimination between looming and other different visual stimuli. In future work, we plan to create a “playground” that contains more realistic moving objects for motion perception experiments.

**V. DISCUSSION**

In biology, both LGMD1 and LGMD2 are believed to be involved in triggering collision avoidance behaviours in locusts; however, their exact roles and actual neural connections are still unknown. A bio-plausible computational model could be a useful tool to unravel the mechanisms behind these fascinating animal behaviours. One good source of such speculation stems from the anatomy and behavioural studies of young locusts [21], in which the LGMD2 grows to a higher state of maturity than the LGMD1 in younger locusts, especially for newborn and first instars. Meanwhile, the “startling” behaviour is observed more frequently in

**TABLE 4.** Actions to be taken by each recognised visual motion.

Recognised Visual Motion	Action Neuron	Speed (cm/s)	Turn angle ( $^{\circ}$ )
safe	forward	6	0
dark loom	U-turn to left	-3	175
bright loom	U-turn to right	-3	-175
translating left	turn left	0	100
translating right	turn right	0	-100
unknown hazard	wait 1 sec	0	0

**FIGURE 18.** The ratio of actions taken by the robot in responding to different visual stimuli, in which each action indicates a recognised motion pattern described in table 4. Each set of experiment is repeated for 50 times.

younger instars when tested with dark looming stimuli. By assuming that LGMD2 plays the dominant role in detecting and reacting to dark approaching objects during locust's young period, with slighter and delayed reactions ("startle") as opposed to stronger and immediate reactions ("hiding"), the proposed visual cue fusion neurons provide a compelling explanation of the biological behavioural studies.

Implementing complex visual algorithms into embedded processors is not an easy task. In the proposed model, the scale of neurons is tailored to fit the resources available, while methods are used to ensure real-time specifically to ensure the computation is almost constant and within the duration of a whole frame. If the requirements of compactness increase, there will be further room for optimisation through the use of a sparse matrix to store and access data, expanding loops into linear calculations, using bit-shift calculations instead of multiplying or dividing, and taking advantage of faster accessing RAM space or using Direct Memory Access (DMA) channels etc.

The parallelism of ViMDNN can be achieved at two levels. Firstly, for each of the presynaptic VMPNs, the layers formed by cell arrays are implemented by low-level operations such as multiplication and accumulation. This internal consistency benefits from parallel processing. Secondly, VMPNs employed in the ViMDNN are involved in minimum data exchange during their processing except for the signal input from the shared  $P$  layer. Therefore, the order of calculation of different VMPNs does not affect the result which is advantageous for multithreading. Meanwhile, unique functions of different VMPNs, such as the selectivity to dark objects and

the direction selectivity, are achieved by parameterised and re-configurable parts of the E-LMGD. This guarantees the ease of accessibility to specific functions. On the platform of Colias-IV, due to the constrain of hardware resources, only four VMPNs have been implemented. When a platform with better support for parallel computing can be utilised, such as Field Programmable Gate Array (FPGA) or even Application Specific Integrated Circuit (ASIC), a ViMDNN with more complex motion selectivity may be conveniently designed.

This work provides a possible solution for low-cost and reliable visual motion perception on micro robot platforms with constrained computational resources. There are two methods to implement the ViMDNN model on other robot platforms. Firstly, the ViMDNN model may be applied via a mainstream microcontroller which has been widely adopted by other robot platforms. Therefore, the model can be easily migrated onto these robots with minor tailoring and modifications. Secondly, thanks to the modular design of the CSU module which contains all the necessary hardware resources for data acquisition and processing, the CSU can be regarded as a "plug and play" module which can be easily attached and integrated within other robot platforms to achieve the desired functions. This is highly advantageous for multi-robot systems such as the bio-inspired swarm robotics research [74].

## VI. CONCLUSION

In this paper, we have proposed a bio-inspired visual motion discrimination neural network (ViMDNN) for micro mobile robots to enable them to effectively react to or interact with the visual events in dynamic visual environments. Equipped with the proposed ViMDNN, a micro mobile robot with a tiny camera and an embedded processor can robustly detect different visual motion cues in real-time at a rate of 30 frames per second and respond with different behaviours accordingly. The compact modularised ViMDNN is realised with a bio-plausible fusion method to integrate multiple motion cues based on E-LGMD. The functionalities and the robustness of the proposed system have been demonstrated and verified with systematic experiments. The proposed embedded visual system shows the possibility of introducing rapid and reliable reactive control strategies to micro-robots in spite of their constrained computational resources. In the future, more integration methods for sequential visual cues under constrained computing resources will be investigated.

## APPENDIX

### A. THE NOTATION OF SIGNALS AND VARIABLES

1) The neuron layers are grid-shaped 2D gray-scale images. Each pixel (neural cell) in most layers is stored as 8-bit unsigned value (0-255), except in the G layer which each cell is 16-bit unsigned value (0-65535). The input image format is YUV and is stored by YUYV format, that every two pixels take 32-bit in total.

2) The processing sequence for each frame is labeled with index ( $f$ ). A neuron layer at frame  $f$  from the processing sequence are labelled in the form of  $P(f)$ ,  $S_{ON}(f)$ ,  $\tilde{G}(f)$  and so on.

3) Individual cell element at coordination ( $x, y$ ) of a certain layer is described in the form such as  $S(x, y, f)$ .

### B. THE BOUNDARY ISSUE OF CONVOLUTION OPERATION

Considering the boundary issue, the calculation of convolution for image border ring is omitted out. For example, we have input layer  $I$  with size  $M \times N$  and a kernel  $K$  with square size  $(2r + 1) \times (2r + 1)$ . Their convolution operation:

$$C = G \otimes K \quad (25)$$

is defined as:

$$C_{(x,y)} = \begin{cases} 0 & \text{for } x=0, \dots, r-1, M-r, \dots, M-1, \\ & y=0, \dots, r-1, N-r, \dots, N-1 \\ \sum_{i=-r}^r \sum_{j=-r}^r I_{(x+i, y+j)} \cdot K_{(i+r, j+r)} & \text{otherwise} \end{cases} \quad (26)$$

### C. THE PRE-S LAYER CALCULATION

B. The operation of  $[x, y]^{\oplus}$  and  $[x, y]^{\ominus}$  in calculating pre-S layers are defined as follows:

$$[x, y]^{\oplus} = \begin{cases} \min(x, y) & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases} \quad (27)$$

$$[x, y]^{\ominus} = \begin{cases} -\max(x, y) & \text{if } x < 0 \text{ and } y < 0 \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

## REFERENCES

- [1] T. Krajnák, F. Majer, L. Halodová, and T. Vintr, "Navigation without localisation: Reliable teach and repeat based on the convergence theorem," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1657–1664.
- [2] F. Gu, H. Gao, J. Liu, Q. Gao, and Z. Ju, "Research on static vision-based target localization for astronaut assistant robots," *IEEE Access*, vol. 7, pp. 128394–128407, 2019.
- [3] Y. Zhou, Q. Fang, K. Zhao, D. Tang, H. Zhou, G. Li, X. Xiang, and T. Hu, "Robust task-oriented markerless extrinsic calibration for robotic pick-and-place scenarios," *IEEE Access*, vol. 7, pp. 127932–127942, 2019.
- [4] T. Cieslewski, S. Choudhary, and D. Scaramuzza, "Data-efficient decentralized visual SLAM," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2466–2473.
- [5] P. Chirarattananon, "A direct optic flow-based strategy for inverse flight altitude estimation with monocular vision and IMU measurements," *Bioinspiration Biomimetics*, vol. 13, no. 3, Mar. 2018, Art. no. 036004.
- [6] J. Whitman, N. Zevallos, M. Travers, and H. Choset, "Snake robot urban search after the 2017 Mexico City earthquake," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot. (SSRR)*, Aug. 2018, pp. 1–6.
- [7] Z. Zhao, S. Lin, D. Zhu, and G. Wen, "Vibration control of a riser-vessel system subject to input backlash and extraneous disturbances," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 3, pp. 516–520, Mar. 2020.
- [8] C. West, F. Arvin, W. Cheah, A. West, S. Watson, M. Giuliani, and B. Lennox, "A debris clearance robot for extreme environments," in *Proc. Annu. Conf. Towards Auto. Robotic Syst.*, London, U.K. Cham, Switzerland: Springer, 2019, pp. 148–159.
- [9] K. Hausen and M. Egelhaaf, "Neural mechanisms of visual course control in insects," *Facets Vis.*, 1989.
- [10] M. V. Srinivasan, R. J. Moore, S. Thurrowgood, D. Soccol, D. Bland, and M. Knight, *Vision and Navigation in Insects, and Applications to Aircraft Guidance*. Cambridge, MA, USA: MIT Press, 2014.
- [11] F. C. Rind, S. Wernitznig, P. Pölt, A. Zankel, D. Gütl, J. Sztarker, and G. Leitinger, "Two identified looming detectors in the locust: Ubiquitous lateral connections among their inputs contribute to selective responses to looming objects," *Sci. Rep.*, vol. 6, no. 1, Dec. 2016.
- [12] A. C. Paulk, J. A. Stacey, T. W. Pearson, G. J. Taylor, R. J. Moore, M. V. Srinivasan, and B. van Swinderen, "Selective attention in the honey-bee optic lobes precedes behavioral choices," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 13, pp. 5006–5011, 2014.
- [13] F. van Breugel and M. H. Dickinson, "The visual control of landing and obstacle avoidance in the fruit fly *Drosophila melanogaster*," *J. Exp. Biol.*, vol. 215, no. 11, pp. 1783–1798, Jun. 2012.
- [14] E. V. Orekhova, E. N. Rostovtseva, V. O. Manyukhina, A. O. Prokofiev, T. S. Obukhova, A. Y. Nikolaeva, J. F. Schneiderman, and T. A. Stroganova, "Spatial suppression in visual motion perception is driven by inhibition: Evidence from MEG gamma oscillations," *NeuroImage*, vol. 213, Jun. 2020, Art. no. 116753.
- [15] C. M. Higgins and V. Pant, "An elaborated model of fly small-target tracking," *Biol. Cybern.*, vol. 91, no. 6, pp. 417–428, Dec. 2004.
- [16] H. Wang, J. Peng, and S. Yue, "A directionally selective small target motion detecting visual neural network in cluttered backgrounds," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1541–1555, Apr. 2020.
- [17] F. Ruffier and N. Franceschini, "Octave, a bioinspired visuo-motor control system for the guidance of micro-air-vehicles," *Proc. SPIE*, vol. 5119, Apr. 2003, pp. 1–12.
- [18] H. Eichner, M. Joesch, B. Schnell, D. F. Reiff, and A. Borst, "Internal structure of the fly elementary motion detector," *Neuron*, vol. 70, no. 6, pp. 1155–1164, Jun. 2011.
- [19] L. D. Liu, R. M. Haefner, and C. C. Pack, "A neural basis for the spatial suppression of visual motion perception," *eLife*, vol. 5, May 2016, Art. no. e16167.
- [20] A. Borst and M. Helmstaedter, "Common circuit design in fly and Mammalian motion vision," *Nature Neurosci.*, vol. 18, no. 8, p. 1067, 2015.
- [21] J. Sztarker and F. C. Rind, "A look into the cockpit of the developing locust: Looming detectors and predator avoidance," *Developmental Neurobiol.*, vol. 74, no. 11, pp. 95–1078, 2014.
- [22] F. C. Rind, "Non-directional, movement sensitive neurones of the locust optic lobe," *J. Comparative Physiol. A*, vol. 161, no. 3, pp. 477–494, 1987.
- [23] P. J. Simmons, J. Sztarker, and F. C. Rind, "Looming detection by identified visual interneurons during larval development of the locust *Locusta migratoria*," *J. Exp. Biol.*, vol. 216, no. 12, pp. 2266–2275, Jun. 2013.
- [24] H. Fotowat, R. R. Harrison, and F. Gabbiani, "Multiplexing of motor information in the discharge of a collision detecting neuron during escape behaviors," *Neuron*, vol. 69, no. 1, pp. 147–158, Jan. 2011.
- [25] Y. Zhu and F. Gabbiani, "Fine and distributed subcellular retinotopy of excitatory inputs to the dendritic tree of a collision-detecting neuron," *J. Neurophysiol.*, vol. 115, no. 6, pp. 3101–3112, Jun. 2016.
- [26] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Comput. Sci. Rev.*, vol. 28, pp. 157–177, May 2018.
- [27] S. Manchanda and S. Sharma, "Analysis of computer vision based techniques for motion detection," in *Proc. 6th Int. Conf.-Cloud Syst. Big Data Eng. (Confluence)*, Jan. 2016, pp. 445–450.
- [28] A. Mitrokhin, Z. Hua, C. Fermuller, and Y. Aloimonos, "Learning visual motion segmentation using event surfaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 14414–14423.
- [29] B. Kalantar, S. B. Mansor, A. A. Halin, H. Z. M. Shafri, and M. Zand, "Multiple moving object detection from UAV videos using trajectories of matched regional adjacency graphs," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5198–5213, Sep. 2017.
- [30] T. Zhao and R. Nevatia, "Tracking multiple humans in complex situations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1208–1221, Sep. 2004.

- [31] K. M. Abughalieh, B. H. Sababha, and N. A. Rawashdeh, "A video-based object detection and tracking system for weight sensitive UAVs," *Multimedia Tools Appl.*, vol. 78, no. 7, pp. 9149–9167, Apr. 2019.
- [32] L. Kaur and U. Mittal, "Moving object recognition and detection using background subtraction," in *Proc. Int. Conf. Commun. Cyber Phys. Eng.* Springer, 2018, pp. 1–10.
- [33] F. H. K. Zaman, M. H. Ali, A. A. Shafie, and Z. I. Rizman, "Efficient human motion detection with adaptive background for vision-based security system," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 3, p. 1026, 2017.
- [34] A. M. Pinto, P. G. Costa, M. V. Correia, A. C. Matos, and A. P. Moreira, "Visual motion perception for mobile robots through dense optical flow fields," *Robot. Auto. Syst.*, vol. 87, pp. 1–14, Jan. 2017.
- [35] B.-D. Wang, J. Song, R. Li, R. Han, G. Zheng, and H. Yang, "A novel particle tracking velocimetry method for complex granular flow field," *Chin. Phys. B*, vol. 29, no. 1, Dec. 2019, Art. no. 014207.
- [36] F. Paredes-Valles, K. Y. W. Schepfer, and G. C. H. E. de Croon, "Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2051–2064, Aug. 2020.
- [37] H. Zhan, R. Garg, C. S. Weerasekera, K. Li, H. Agarwal, and I. M. Reid, "Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 340–349.
- [38] Y. Chen, J. Wang, B. Zhu, M. Tang, and H. Lu, "Pixelwise deep sequence learning for moving object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 9, pp. 2567–2579, Sep. 2017.
- [39] A. Mitrokhin, C. Fermüller, C. Parameshwara, and Y. Aloimonos, "Event-based moving object detection and tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1–9.
- [40] S. Yue and F. C. Rind, "A synthetic vision system using directionally selective motion detectors to recognize collision," *Artif. Life*, vol. 13, no. 2, pp. 93–122, Apr. 2007.
- [41] M. Blanchard, F. C. Rind, and P. F. M. J. Verschure, "Collision avoidance using a model of the locust LGMD neuron," *Robot. Auto. Syst.*, vol. 30, nos. 1–2, pp. 17–38, Jan. 2000.
- [42] C. Hu, F. Arvin, C. Xiong, and S. Yue, "Bio-inspired embedded vision system for autonomous micro-robots: The LGMD case," *IEEE Trans. Cogn. Develop. Syst.*, vol. 9, no. 3, pp. 241–254, Sep. 2017.
- [43] S. Yue and F. C. Rind, "Collision detection in complex dynamic scenes using an LGMD-based visual neural network with feature enhancement," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 705–716, May 2006.
- [44] Q. Fu, S. Yue, and C. Hu, "Bio-inspired collision detector with enhanced selectivity for ground robotic vision system," *Trans. Neural Netw.*, vol. 17, no. 3, pp. 705–716, 2016.
- [45] H. Meng, K. Appiah, S. Yue, A. Hunter, M. Hobden, N. Priestley, P. Hobden, and C. Pettit, "A modified model for the lobula giant movement detector and its FPGA implementation," *Comput. Vis. Image Understand.*, vol. 114, no. 11, pp. 1238–1247, Nov. 2010.
- [46] A. Krejan and A. Trost, "LGMD-based bio-inspired algorithm for detecting risk of collision of a road vehicle," in *Proc. Int. Symp. Image Signal Process. Anal.*, 2011, pp. 319–324.
- [47] 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 Trans. Cybern.*, early access, Dec. 4, 2019, doi: [10.1109/TCYB.2019.2946090](https://doi.org/10.1109/TCYB.2019.2946090).
- [48] J. Zhao, C. Hu, C. Zhang, Z. Wang, and S. Yue, "A bio-inspired collision detector for small quadcopter," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7.
- [49] L. Salt, D. Howard, G. Indiveri, and Y. Sandamirskaya, "Parameter optimization and learning in a spiking neural network for UAV obstacle avoidance targeting neuromorphic processors," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 14, 2019, doi: [10.1109/TNNLS.2019.2941506](https://doi.org/10.1109/TNNLS.2019.2941506).
- [50] B. Hassenstein and W. Reichardt, "Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenbewertung bei der bewegungsperzeption des rüsselkäfers chlorophanus," *Zeitschrift für Naturforschung B*, vol. 11, nos. 9–10, pp. 513–524, 1956.
- [51] M. S. Maisak, J. Haag, G. Ammer, E. Serbe, M. Meier, A. Leonhardt, T. Schilling, A. Bahl, G. M. Rubin, A. Nern, B. J. Dickson, D. F. Reiff, E. Hopp, and A. Borst, "A directional tuning map of *Drosophila* elementary motion detectors," *Nature*, vol. 500, no. 7461, p. 212, 2013.
- [52] M. F. Keleş, B. J. Hardcastle, C. Städele, Q. Xiao, and M. A. Frye, "Inhibitory interactions and columnar inputs to an object motion detector in *drosophila*," *Cell Rep.*, vol. 30, no. 7, pp. 2115–2124, 2020.
- [53] Q. Fu and S. Yue, "Mimicking fly motion tracking and fixation behaviors with a hybrid visual neural network," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2017, pp. 1636–1641.
- [54] 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," *J. Neural Eng.*, vol. 14, no. 4, Aug. 2017, Art. no. 046030.
- [55] M. Srinivasan and S. Zhang, "Visual control of honeybee flight," in *Orientation and Communication in Arthropods*. Springer, 1997, pp. 95–113.
- [56] M. V. Srinivasan, R. J. Moore, S. Thurrowgood, D. Soccol, and D. Bland, *From Biology to Engineering: Insect Vision and Applications to Robotics*. Springer, 2012.
- [57] Y. Yamawaki, K. Uno, R. Ikeda, and Y. Toh, "Coordinated movements of the head and body during orienting behaviour in the praying mantis *tenodera aridifolia*," *J. Insect Physiol.*, vol. 57, no. 7, pp. 1010–1016, Jul. 2011.
- [58] N. S. Szczecinski, J. P. Martin, D. J. Bertsch, R. E. Ritzmann, and R. D. Quinn, "Neuromechanical model of praying mantis explores the role of descending commands in pre-strike pivots," *Bioinspiration Biomimetics*, vol. 10, no. 6, Nov. 2015, Art. no. 065005.
- [59] F. C. Rind, L. N. Jones, G. Tarawneh, and J. Read, "Target tracking behaviour of the praying mantis *sphodromantis lineola* (Linnaeus) is driven by looming-type motion-detectors," *bioRxiv*, to be published, doi: [10.1101/2020.06.02.129684](https://doi.org/10.1101/2020.06.02.129684).
- [60] S. Bermudez I Badia, P. Pyk, and P. F. M. J. Verschure, "A fly-locust based neuronal control system applied to an unmanned aerial vehicle: The invertebrate neuronal principles for course stabilization, altitude control and collision avoidance," *Int. J. Robot. Res.*, vol. 26, no. 7, pp. 759–772, Jul. 2007.
- [61] S. Yue and F. C. Rind, "Postsynaptic organisations of directional selective visual neural networks for collision detection," *Neurocomputing*, vol. 103, pp. 50–62, Mar. 2013.
- [62] B. Hu, S. Yue, and Z. Zhang, "A rotational motion perception neural network based on asymmetric spatiotemporal visual information processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2803–2821, Nov. 2017.
- [63] S. Yue and F. C. Rind, "Redundant neural vision systems—Competing for collision recognition roles," *IEEE Trans. Auto. Mental Develop.*, vol. 5, no. 2, pp. 173–186, Jun. 2013.
- [64] Q. Fu, C. Hu, T. Liu, and S. Yue, "Collision selective LGMDs neuron models research benefits from a vision-based autonomous micro robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 3996–4002.
- [65] B. Hu and Z. Zhang, "Bio-plausible visual neural network for spatio-temporally spiral motion perception," *Neurocomputing*, vol. 310, pp. 96–114, Oct. 2018.
- [66] B. Hassenstein and R. Hustert, "Hiding responses of locusts to approaching objects," *J. Exp. Biol.*, vol. 202, no. 12, pp. 1701–1710, 1999.
- [67] 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. 108896–108912, 2020.
- [68] L. F. Tammero and M. H. Dickinson, "The influence of visual landscape on the free flight behavior of the fruit fly *drosophila melanogaster*," *J. Exp. Biol.*, vol. 205, no. 3, pp. 327–343, 2002.
- [69] J.-C. Zufferey, A. Beyeler, and D. Floreano, "Vision-based navigation from wheels to wings," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, vol. 3, Oct. 2003, pp. 2968–2973.
- [70] J.-C. Zufferey and D. Floreano, "Fly-inspired visual steering of an ultralight indoor aircraft," *IEEE Trans. Robot.*, vol. 22, no. 1, pp. 137–146, Feb. 2006.
- [71] A. Borst and S. Bahde, "Visual information processing in the fly's landing system," *J. Comparative Physiol. A*, vol. 163, no. 2, pp. 167–173, 1988.
- [72] H. Cheng, F. Qinbing, and Y. Shigang, "Colias IV: The affordable micro robot platform with bio-inspired vision," in *Proc. 19th Towards Auto. Robotic Syst. (TAROS)*, 2018, pp. 197–208.
- [73] P. Lightbody, T. Krajník, and M. Hanheide, "An efficient visual fiducial localisation system," *ACM SIGAPP Appl. Comput. Rev.*, vol. 17, no. 3, pp. 28–37, Nov. 2017.
- [74] J. Hu, P. Bhowmick, F. Arvin, A. Lanzon, and B. Lennox, "Cooperative control of heterogeneous connected vehicle platoons: An adaptive leader-following approach," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 977–984, Apr. 2020.





**CHENG HU** (Member, IEEE) received the B.Eng. degree from the College of Opto-Electronics, Chongqing University, Chongqing, China, in 2013, and the Ph.D. degree from the University of Lincoln, Lincoln, U.K., in 2017.

He is currently a Postdoctoral Research Fellow of Guangzhou University, Guangzhou, and an Honorary Postdoctoral Research Fellow of the University of Lincoln. His current research interests include bio-inspired algorithms, bio-robotics, swarm robotics, sensors, and image processing.

Dr. Hu was a recipient of the Marie Curie Fellowship, involved in the EU-FP7 Projects EYE2E (269118) and LIVECODE (295151), and the EU Horizon 2020 Projects STEP2DYNA (691154) and ULTRACEPT (778062).



**CAIHUA XIONG** (Member, IEEE) received the Ph.D. degree in mechanical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1998.

From 1999 to 2003, he was a Postdoctoral Fellow of the City University of Hong Kong, Hong Kong, and The Chinese University of Hong Kong, Hong Kong, and a Research Scientist with the Worcester Polytechnic Institute, Worcester, MA, USA. Since 2008, he has been a Chang

Jiang Professor with the HUST, where he is also the Director of the Institute of Rehabilitation and Medical Robotics. His current research interests include biomechatronic prostheses, rehabilitation robotics, and robot motion planning and control.



**JIGEN PENG** received the B.Sc. degree from Jiangxi University, Nanchang, China, in 1989, and the M.Sc. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1992 and 1998, respectively.

He was with Xi'an Jiaotong University and the City University of Hong Kong, Hong Kong, until 2017. He is currently a Professor with the School of Mathematics and Information Science, Guangzhou University, Guangzhou, China. His

current research interests include nonlinear functional analysis and its applications, machine-learning theory, and sparse information processing.



**SHIGANG YUE** (Senior Member, IEEE) received the B.Eng. degree from Qingdao Technological University, Qingdao, China, in 1988, and the M.Sc. and Ph.D. degrees from the Beijing University of Technology (BJUT), Beijing, China, in 1993 and 1996, respectively. He was a Lecturer with the BJUT from 1996 to 1998 and an Associate Professor from 1998 to 1999. He was an Alexander von Humboldt Research Fellow of the University of Kaiserslautern, Kaiserslautern, Germany, from

2000 to 2001. He held research positions at the University of Cambridge, Cambridge, U.K., Newcastle University, Newcastle upon Tyne, U.K., and University College London, London, U.K., from 2002 to 2007. In 2007, he joined the University of Lincoln, Lincoln, U.K., as a Senior Lecturer, and was promoted to a Professor in 2012, where he is currently a Professor (part-time) with the School of Computer Science. He also holds a Professorship with the Machine Life and Intelligence Research Centre, Guangzhou University, Guangzhou, China, in collaboration with Prof. J. Peng. His current research interests include artificial intelligence, computer vision, robotics, intelligent transportation, brains, and neuroscience. He is a member of the International Neural Network Society and the International Society of Bionic Engineering.

...