



UNIVERSITY OF
LINCOLN

**Bio-inspired Collision Detection
with Motion Cues Enhancement in
Dim Light Environments**

Fang Lei

Doctor of Philosophy

2022

School of Computer Science

University of Lincoln

Bio-inspired Collision Detection with Motion Cues Enhancement in Dim Light Environments

Fang Lei

School of Computer Science

University of Lincoln

A thesis submitted in partial fulfilment of the requirements of the University of
Lincoln for the degree of Doctor of Philosophy

Supervisor

Professor Shigang Yue

May 2022

Abstract

Detecting looming objects robustly and timely is a huge challenge for artificial vision systems in complex natural scenes, including dim light scenes. Insects have evolved remarkable capacities in collision detection despite their tiny eyes and brains. The locusts' LGMD1 neuron shows strong looming-sensitive property for both light and dark objects, which is a source of inspiration for developing collision detection systems. Furthermore, specialized visual processing strategies in nocturnal animals' brains can provide inspiration for detecting faint motion like dim-light collision detection when challenged with low light conditions. This research aims to explore the LGMD1 based collision detection methods, adaptive low-light image enhancement methods, biologically-inspired solutions for enhancing faint motion cues as well as collision detection methods in low light conditions. The major contributions are summarized as follows.

A new visual neural system model (LGMD1) is developed, which applies a neural competition mechanism within a framework of separated ON and OFF pathways to shut off the translating response. The competition-based approach responds vigorously to monotonous ON/OFF responses resulting from a looming object. However, it does not respond to paired ON-OFF responses that result from a translating object, thereby enhancing collision selectivity. Moreover, a complementary denoising mechanism ensures reliable collision detection. To verify the effectiveness of the model, we have conducted systematic comparative experiments on synthetic and real datasets. The results show that our method exhibits more accurate discrimination between looming and translational events—the looming motion can be correctly detected. It also demonstrates that the proposed model is more robust than comparative models.

A framework is proposed for adaptively enhancing low-light images, which im-

plements the processing of dark adaptation with proper adaptation parameters in R, G and B channels separately. Specifically, the dark adaptation processing consists of a series of canonical neural computations, including the power law adaptation, divisive normalization and adaptive rescaling operations. Experimental results show that the proposed bioinspired dark adaptation framework is more efficient and can better preserve the naturalness of the image compared with several representative low light image enhancement methods.

A dim-light motion cues enhancement (DLMCE) model is designed for extracting extremely faint motion cues. This model integrates dark-adaptation, spatio-temporal constraint and neural summation mechanisms, which are achieved with canonical neural computations and neural summation in temporal and spatial domains, to enhance faint motion cues. With the DLMCE model, the image intensity and contrast are first increased by the dark adaptation processing, then the strong motion cues are extracted by the spatio-temporal constraint strategy, and these motion cues are further enhanced by neural summation mechanisms. Experimental results have demonstrated that the presented DLMCE model outperforms the existing methods for dim-light motion cues enhancement, and faint motion cues can be successfully detected in consecutive frames efficiently. As demonstrated in the experiments, the proposed DLMCE model provides a robust and effective solution for autonomous systems in detecting moving objects under low light conditions.

A bio-inspired collision detection model is developed for detecting looming objects in dim light environments. The model combines the DLMCE model with the classical four-layered LGMD1 model to detect dimly illuminated approaching objects. To verify the effectiveness of the model, we have conducted comparative experiments on real looming datasets. The results have demonstrated that the proposed bio-inspired collision detection model can correctly recognize looming objects under low light conditions since the DLMCE model enhances the faint looming cues.

Declaration

I, Fang Lei, declare that this thesis describes an original research carried out on my own. It has not been previously submitted to any university for the award of any degree. Where I have quoted from the work of others, the source is always given.

Acknowledgements

First and foremost, I would like to thank my first supervisor, Professor Shigang Yue. He is so tolerant and patient in letting me explore my interest fields during my PhD studying period. He always encouraged me when I encountered problems in my research area, and he also taught me how to effectively carry out research work and write academic paper. He gives me plenty of chances without expecting anything in return. He will always be a role model in my academic career and I have learned many things from him, especially for his effective communication skills, hard-working attitude and pursuing high quality research.

Professor Vassilis Cutsuridis was my second supervisor at University of Lincoln, and I am grateful to him for his help and support. I would also like to thank Professor Mei Liu and Professor Zhiping Peng who helped me a lot so that I had the chance to study at University of Lincoln. I am also very grateful to my colleagues Xiaowen Liao, Yicheng Zhang and Li Yan, who offered me valuable help and great support and have also been my schoolmates.

At University of Lincoln, there is a lot of brilliant and enthusiastic colleagues who have provided me valuable help and discussions. I would also like to thank the past and present members in the Computational Intelligent Lab (CIL) who have brought me great convenience on daily life and academic research. Particular thanks to Hongxin Wang and Jiannan Zhao who provided selfless help and creative discussions.

Then, I appreciate the financial support from School of Computer Science at University of Lincoln and EU Horizon 2020 project STEP2DYNA (691154) and ULTRA-CEPT (778062), as well as the possible financial support provided by the Joint Research Lab between Guangdong University of Petrochemical Technology and University of Lincoln.

Finally, my deep and sincere gratitude to my family, especially my husband Gao Liu who gave me great support on deciding to start my PhD study and supported the whole home when I was studying abroad. And I also want to say sorry to my son for the absence during my abroad study.

List of Main Publications

- [1] **F. Lei**, Z. Peng, V. Cutsuridis, M. Liu, Y. Zhang and S. Yue, “Competition between ON and OFF Neural Pathways Enhancing Collision Selectivity,” in *Proc. Int. Jt. Conf. Neural. Netw. (IJCNN)*, Jul. 2020, pp. 1-8.
- [2] **F. Lei**, Z. Peng, M. Liu, J. Peng, V. Cutsuridis and S. Yue, “A Robust Visual System for Looming Cue Detection Against Translating Motion,” *IEEE Trans. Neural Netw. Learn. Syst. (T-NNLS)*, Feb. 2022, published online.
- [3] **F. Lei**, “A Bio-inspired Dark Adaptation Framework for Low-light Image Enhancement,” *Proc. Int. Jt. Conf. Neural. Netw. (IJCNN)*, 2022, accepted.
- [4] **F. Lei**, H. Wang, J. Peng, S. Yue, “A Bio-inspired Visual System for Dim-light Motion Cues Enhancement in Low Light Conditions,” *IEEE Trans. Neural Netw. Learn. Syst. (T-NNLS)*, 2022, under review.
- [5] **F. Lei**, S. Yue, “A Bio-inspired Dim-light Collision Detection Model,” *Proc. Int. Jt. Conf. Neural. Netw. (IJCNN)*, 2023, to be submitted.

Contents

Abstract	I
Declaration	III
Acknowledgements	IV
List of Publications	VI
List of Figures	IX
List of Abbreviations	X
1 Introduction	1
1.1 Motivation	1
1.2 Organization of the Thesis	4
2 Related Work	6
2.1 Biologically Inspired Collision Detection Methods	7
2.1.1 Directionally Selective Neurons	7
2.1.2 LGMD1 based Neural Network Models	10
2.2 Traditional Collision Detection Methods	13
2.2.1 Optical-flow based methods	16
2.2.2 Feature based methods	18
2.3 Neural Mechanisms in Nocturnal Vision	20
2.3.1 Photoreceptors' Dark Adaptation	20
2.3.2 Canonical Neural Computations	24
2.3.3 Neural Summation	25
2.4 Low-light Image Enhancement Methods	28
2.4.1 Histogram-based Methods	29
2.4.2 Retinex-based Methods	30
2.4.3 Gradient-based Methods	31
2.4.4 Deep Learning-based Methods	31
2.5 Motion Cues Detection Methods	32
2.5.1 Temporal Differencing	32
2.5.2 Optical Flow	33
2.5.3 Motion Detectors	34
3 LGMD1- Looming Cue Detection Against Translating Motion	37
3.1 Formulation of the LGMD1 Model	39
3.1.1 Retina to Lamina layer	40
3.1.2 Lamina to Medulla layer	41

3.1.3	Medulla to Lobula layer	44
3.1.4	Spiking Mechanism	46
3.1.5	The Feed Forward Inhibition (FFI)	47
3.1.6	Parameters of the System	48
3.2	Experimental Results	48
3.2.1	Experimental Setup	48
3.2.2	ON/OFF Responses to Looming and Translating Stimuli	50
3.2.3	Effectiveness of the Denoising Mechanism	52
3.2.4	Effectiveness of Suppression on Saccadic Eye Movements	54
3.2.5	Comparison on Synthetic and Real Datasets	59
3.2.6	Comparison with Conventional Methods	65
3.3	Chapter Summary	67
4	Dark Adaptation- Low Light Image Enhancement	73
4.1	Bio-inspired Dark Adaptation Framework	75
4.1.1	Power Law Adaptation	76
4.1.2	Divisive Normalization	79
4.1.3	Adaptive Rescaling	81
4.2	Experimental Results	83
4.2.1	Comparison on Real Low-Light Dataset	84
4.3	Discussion	88
4.4	Chapter Summary	88
5	DLMCE- Dim-light Motion Cues Enhancement	90
5.1	Formulation of the DLMCE Model	92
5.1.1	Retina	93
5.1.2	Lamina	96
5.1.3	Parameters of the System	99
5.2	Experimental Results	100
5.2.1	Experimental Setup	100
5.2.2	Effectiveness of the Dark Adaptation	102
5.2.3	Effectiveness of the Motion Cues Enhancement	105
5.2.4	Comparison on Synthetic and Real Datasets	107
5.3	Chapter Summary	117
6	Bio-inspired Collision Detection in Dim Light Environments	118
6.1	Formulation of the Bio-inspired Dim-light Collision Detection Model	120
6.1.1	DLMCE layer	120
6.1.2	P_M layer	125
6.1.3	E_M/I_M layer	125
6.1.4	S_M layer	126
6.1.5	LGMD1 cell	126
6.1.6	Spiking Mechanism	127
6.1.7	The Feed Forward Inhibition (FFI)	127
6.1.8	Parameters of the System	128
6.2	Experimental Results	128
6.2.1	Experimental Setup	128
6.2.2	Effectiveness of Looming Motion Detection	129
6.2.3	Comparison with Comparative LGMD1 Models	129
6.3	Chapter Summary	133

7	Research Contributions and Future Work	136
7.1	Research Contributions	136
7.2	Future Work	137
	Bibliography	138

List of Figures

- 1.1 (a) The LGMD neuron's dendritic tree, which consist of three distinct subfields (A–C). This figure is adapted from [1]. (b) A looming black ball under dim light environment. 2
- 2.1 Schematic illustration of the EMD process in four steps. The figure is adapted from [2]. 8
- 2.2 Schematic illustration of four classical EMD models. (a) Hassenstein-Reichardt-(half)-detector: preferred-direction (PD) enhancement. (b) Barlow-Levick-detector: null-direction (ND) suppression. (c) (full) Hassenstein-Reichardt-detector. (d) HR/BL-detector: PD-enhancement + ND-suppression. The figure is adapted from [3]. 8
- 2.3 Schematic illustration of the locust visual system: Presynaptic neuropile layers of the LGMD1 neuron and its postsynaptic target neuron, the descending contralateral movement detector (DCMD). The figure is adapted from [4]. 10
- 2.4 Schematic of LGMD1 visual neural network proposed by Rind and Bramwell [5]. The model consists of four layers, including a LGMD1 cell and a FFI cell. The input organization of the basic retinotopic unit of the network is shown in red and labeled. The outputs of these units were summed by the LGMD unit (blue). In layer 1, images were mapped onto the photoreceptors. Each photoreceptor (P unit, layer 1) in the model responded with a brief excitation to a change in level of illumination (graph P). In layer 2, this excitation was passed to three units: E, I and F. When excited, E passed excitation to a layer 3, S unit in the same retinotopic position. I passed delayed inhibition laterally to the nearest S units in the layer 3. F fed inhibition forward and by-passed layer 3. The E and I inputs were summed linearly by each S unit in layer 3 until a threshold level of excitation was reached and a spike was produced. Layer 4 of the model consisted of a single LGMD unit that summed excitation from all active S units and inhibition from the F unit. In each layer proximity to the central retinotopic unit is indicated by the shade of red (dark red is the nearest). The figure is adapted from [4]. 12
- 2.5 The schematic illustration of the LGMD1 based neural network and the asymmetric lateral inhibition based TSNN for colliding objects detection. (a) The LGMD1 based neural network. (b) The asymmetric lateral inhibition based TSNN. The figure is adapted from [6]. 13

- 2.6 The LGMD1 model proposed by Yue and Rind [7]. There are five groups of cells and two single cells: Photoreceptor cells (P); excitatory and inhibitory cells (E and I); summing cells (S); grouping cells (G); the LGMD cell; and the feed forward inhibition cell (FFI). The input of the P cells is the luminance change. Lateral inhibition is indicated with dotted lines and has one frame delay. Excitation is indicated with black lines and has no delay. The FFI also has one frame delay. The input to FFI is the luminance changes from photoreceptor cells. 14
- 2.7 The LGMD1 model proposed by Fu et al. [8]. The proposed LGMD1 model processes signals in the ON and OFF pathways without bias, whilst a new SFA mechanism is modeled. The dashed lines indicate transmissions of delayed neural signals. 14
- 2.8 The neural competition based LGMD1 model. The model is composed of five layers (P, E, I, S, G) and three cells (FFIon, FFIOff, LGMD1); signals are split into ON (red-arrows) and OFF (blue-arrows) pathways each with four layers (E, I, S, G); excitatory signals make a neural competition in LGMD1 neuron. Note that all transmitted signals with time delays are indicated with dashed lines. The figure is adapted from [9] 15
- 2.9 (a) Apposition compound eyes. There are nine ommatidia here. The photoreceptors receive light exclusively from the single facet lens of their own ommatidium. (b) Superposition compound eyes. The lenses and photoreceptors are separated by a wide optically clear region known as the clear zone (cz). The lenses allow light from many hundreds of facet lenses to be focused onto single photoreceptors in the retina, thus dramatically increasing light capture. The figure is adapted from [10]. 21
- 2.10 Responses to single photons (red and blue arrowheads) recorded from photoreceptors in (a) nocturnal sweat bee (*Megalopta genalis*) and (b) diurnal sweat bee (*Lasioglossum leucozonium*). The figure is adapted from [11]. 22
- 2.11 The voltage responses of locust photoreceptors. a, b The saturated response of a dark adapted cell recorded on fast and slow time base. c, d Responses of a light adapted cell to an increment above background intensity, and to a brief extinction of the background. e Dark adapted and (f, g) two light adapted responses of different amplitudes. The figure is adapted from [12]. 23
- 2.12 Normalization in the retina. (a) Intensity normalization. The multiple arrows indicate light intensities from multiple locations. (b) Responses of intensity normalization, data from [13]. (c) Contrast normalization. (d) Responses of contrast normalization, data from [14]. The figure is adapted from [15]. 26
- 2.13 Dendritic fields of the short visual fibres (svf 1–3) and first-order interneurons or L-fibres (L1-4) within the first optic ganglion (lamina) of the nocturnal bee *Megalopta genalis*. The black circle indicates the parental cartridge of the fibre, and the neighbouring cartridges reached by its dendrites are shown by grey circles. The lamina has three layers (A, B, and C). Note that v and d indicate ventral and dorsal respectively. The figure is adapted from [16]. 27

- 2.14 Schematic of the basic summation model. The input channel receptive field is modeled as a Gaussian of half-width $\Delta\rho$, and the output channel receptive field is also Gaussian with a half-width $\Delta\rho_T$. In the process of spatiotemporal summation, the spatial summation function has a half-width of $\Delta\rho_p$ degrees and the motion blurring function has a half-width $v\Delta_t$ degrees. The function for both the spatial and temporal summation is Gaussian. The figure is adapted from [17]. 28
- 2.15 Motion cues on the result of frame differencing. The figure is adapted from [18]. 33
- 2.16 Motion boundaries of moving objects by the computation of optical flow. (a) Two input frames. (b) Optical flow \vec{f}_p . (c) Motion boundaries, based on the magnitude of the gradient of the optical flow b_p^m . (d) Motion boundaries, based on difference in direction between a pixel and its neighbors b_p^θ . (e) Combined motion boundaries b_p . (f) Final, binary motion boundaries after thresholding, overlaid on the first frame. The figure is adapted from [18]. 35
- 2.17 Motion cues resulting from specific layers of LGMD1 and TSNNs. a. LGMD1 neural network model. b. TSNN neural network model, which includes four DSNs (L-Left, R-Right, U-Up, D-Down). The figure is adapted from [6]. 36
- 3.1 The proposed LGMD1 model is composed of six layers of cells (P, E, I, D_E, D_I, S) and three single cells (FFI_{on}, FFI_{off}, LGMD1). Signals are split into ON (red-arrows) and OFF (blue-arrows) channels each with five layers (E, I, D_E, D_I and S); the D_E and D_I layers are responsible for filtering out isolated background excitation. The dashed lines indicate transmission of delayed neural signals. Excitatory signals from the two channels create neural competition in the LGMD1 neuron. 39
- 3.2 Schematic illustration of the three stages of signal processing in the LGMD1 model. Retina to lamina layer: the photoreceptor (P) cell captures changes in pixel-wise luminance (L); each P-unit feeds both an ON and an OFF channel. Lamina to medulla layer: two types of visual stimuli are processed separately in ON and OFF channels. This includes an asymmetric mechanism (rectification), a denoising mechanism (see Fig. 3.3) and a lateral inhibition mechanism (see Fig. 3.4). Medulla to lobula layer: the LGMD1 neuron collects the winner excitatory signals arising from the neural competition between ON and OFF pathways; the delayed feed forward inhibition (FFI) signals from the ON and OFF responses inhibit the initial response of the model to movement or sudden great changes in luminance in the visual field. 40
- 3.3 Schematic illustration of the denoising mechanism. The red cells in E/I layer represent the receptive field of the cell (red) in the D_E/D_I layer. Two processing stages are implemented between the two layers, including the computation of the passing coefficient (the black arrows) and the transfer of the expected excitation or inhibition signal (the blue arrow). 42

- 3.4 Schematic illustration of the lateral inhibition mechanism. (a) The excitation (D_E) unit and the neighboring inhibition (D_I) unit from the D layer are summed to the summation (S) unit in the same retinotopic position; Excitation and inhibition decay exponentially with different coefficients, and the decay processes are described in [5]; Delays at excitatory connections are set to 0 milliseconds, whereas delays at inhibitory connections are set to vary from several to tens of milliseconds accordingly (one frame delay). (b) Connection weight matrix of lateral inhibition. 45
- 3.5 Schematic illustration of the neural competition between ON and OFF channels. S_{ON} and S_{OFF} are effective outputs (nonzero value) from the S layer participating in competition. The winner excitations are passed to the LGMD1 neuron. 45
- 3.6 Outputs of the P layer when the model is tested with looming and translating stimuli. The left column presents looming or translational motion, including dark objects moving on light backgrounds as well as light objects moving on dark backgrounds. The middle column indicates outputs of the P layer of a specified pixel in the t axis. The right column denotes outputs of horizontal line pixels in the image plane at a fixed time point in the x axis. (a) and (b) show outputs of testing looming stimuli. (c) and (d) show outputs of testing translating stimuli. 51
- 3.7 Responses to looming and translating stimuli respectively, including responding to ON and OFF channels, and the normalized membrane potential (NMP) of the LGMD1 neuron. The blue curve shows excitations from the ON and OFF pathways. The red curve denotes the normalized membrane potential (NMP) of LGMD1. The green curve represents the height of the square for looming stimuli and the position in the horizontal axis for translating stimuli independently. (a) A, B: approaching squares. C, E: response of looming stimuli at constant speed; D, F: responses to looming stimuli at loom-like increasing speed. (b) A, B: translationally moving squares. C, E: responses to translating stimuli at constant speed; D, F: responses to translating stimuli at loom-like increasing speed. Note that the spiking threshold is set at 0.7 in accordance with [8]. 53
- 3.8 Schematic illustration of early visual processing in the proposed model. The input image was contaminated by random noise and it was passed into the P layer for extraction and denoising of motion stimuli. The processing stages were then implemented in the P, E/I, and D_E/D_I layers. 55
- 3.9 Neural responses of the proposed LGMD1 model that adopts various processing strategies. The visual stimuli tested consist of the image sequence with different levels of random noise ($k_n = 0, 10, 20, 30, 40, 50$). (a) The sample frames of the image sequence to test the model. (b) The model in the absence of either the GD processing or denoising mechanisms (without G & D layers). (c) The model implementing GD processing (with the G layer). (d) The model implementing the denoising mechanism (with the D layer). 56

- 3.10 Neural responses to sinusoidal gratings with the spiking threshold set at **0.7**. (a) The example gratings as input. (b) The model is challenged by a series of gratings with different temporal frequencies. (c) The model is challenged by a series of gratings with different spatial frequencies. 57
- 3.11 (a) Three video sequences of moving backgrounds are presented: top row (moving background 1), middle row (moving background 2), and bottom row (moving background 3). Each video sequence is presented with three frames; the frame number is indicated under each image. (b) Neural responses to three moving backgrounds. 58
- 3.12 Example images of simulated motion events; neural responses of models to these moving stimuli; symbols *, o, and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. (a)-(c) Output curves of looming versus translational events at different speeds. (d)-(e) Output curves of looming versus translational events under different contrasts. **The red curves show that the proposed model responds selectively to looming objects at various speeds/contrasts but not to translating objects, while other models show no preference between the looming and the translating objects.** 60
- 3.13 Example images of real motion events with simple backgrounds; neural responses of models to movements by objects; symbols *, o, and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. The object approaches the camera: (a) the black ball, (d) the white ball. A translating object moves from left to right in front of the camera: (b) the black ball, (e) the white ball. Two translating objects move from the center to left and right separately: (c) two black balls, (f) two white balls. **The red curves show the proposed model prefers to respond to looming objects rather than translating ones against clean backgrounds.** 62
- 3.14 Example images of real motion events with complex backgrounds; neural responses of models to movements by objects; symbols *, o, and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. The object approaches the camera: (a) the blue toy car, (d) the white toy car. A translating object moves from left to right in front of the camera: (b) the black ball, (e) the white ball. Two translating objects move from the center to left and right separately: (c) the black ball and the white ball, (f) the blue toy car and the white toy car. **The red curves show the proposed model prefers to respond to looming objects rather than to translating ones against complex backgrounds.** 63

- 3.15 Example images of real motion events with complex backgrounds; neural responses of models to movements by the camera; symbols *, o, and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. The camera approaches the object: (a) the blue toy car, (d) the white toy car. The camera turns from right to left in front of objects: (b) the blue toy car, (e) the white car, (c) the black ball and the white ball, (f) the blue toy car and the white toy car. **The red curves show the proposed model prefers to respond to looming objects rather than other translating cues with the camera in motion.** 64
- 3.16 Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 1** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E , which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into NP, NP, NP and TN from column one to column four. (d) The outputs of FD-S, which are categorized into NP, NP, TN and TN from column one to column four. (e) The outputs of FD-A, which are categorized into NP, NP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that untextured object against white background.** 68
- 3.17 Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 2** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E , which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into TP, TP, TN and TN from column one to column four. (d) The outputs of FD-S, which are categorized into TP, TP, FP and TN from column one to column four. (e) The outputs of FD-A, which are categorized into TP, TP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that textured object against white background.** 69

- 3.18 Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 3** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E, which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into FN, TP, TN and TN from column one to column four. (d) The outputs of FD-S, which are categorized into TP, TP, FP and TN from column one to column four. (e) The outputs of FD-A, which are categorized into FN, TP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that untextured object against complex background.** 70
- 3.19 Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 4** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E, which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into FN, TP, TN and TN from column one to column four. (d) The outputs of FD-S, which are categorized into FN, TP, TN and FP from column one to column four. (e) The outputs of FD-A, which are categorized into FN, TP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that textured object against complex background.** 71
- 4.1 Proposed dark adaptation framework for low light image enhancement. The red (R), green (G), and blue (B) components of the input image are processed with the dark adaptation in three separate channels. Note that each channel has a different adaptation parameter. 75
- 4.2 Schematic illustration of the proposed dark adaptation. There are N cells that correspond to N pixels in the input image, denoted by $I_1 \sim I_N$. n denotes the sensation parameter, and its value depends on the wavelength of perceived light. I_i and I'_i indicate the i th cell and its enhanced output after the dark adaptation processing. For clear illustration, we only give one cell's enhanced result. 76
- 4.3 The perceived intensity ψ for different values of the adaptation parameter n (refer to Eq. (4.1), $a = 1$, $I_0 = 0$). 77
- 4.4 The same scenes are captured by a GoPro camera under different illumination conditions. (a) Reference image (normal illumination). (b) Dark image (low illumination). 78

- 4.5 The divisive normalization operation for different value of the additive term I_{m_R} , I_{m_G} and I_{m_B} (refer to Eq. 4.4). Here, we set the average intensity in each color channel (R, G, and B) with the same value, $I_{m_R} = I_{m_G} = I_{m_B} = I_m$. (a) $I_m = 100$. (b) $I_m = 50$. (c) $I_m = 10$. 82
- 4.6 The results of adaptive rescaling operation with respect to the adaptation parameters $n_R = 0.6$, $n_G = 0.5$ and $n_B = 0.4$ (refer to Eq. 4.5). 83
- 4.7 Sample images in the low-light image dataset. From the first column to the third column, images are captured under different levels of illumination. The scenes recorded by the images in each row are the same. Specifically, scene1 and scene2 are static images while scene3 and scene4 are images from video clips. Note that images in the first column are captured under higher illumination, which is used in the evaluation experiments as the reference images for the dark and very dark scenes in the second and the third columns. 86
- 4.8 Visual comparison among the competitors on the low-light image dataset. 87
- 5.1 The proposed DLMCE model is composed of six layers of cells (L, P, E, I, S, M) from the retina to the lamina. The input to the L cell corresponds to the pixel's luminance. The P cell perceives the changes in luminance and passes them to the E/I cell. The S cell accumulates signals from the E and I cells, including lateral excitation and self-inhibition. Note that the signal without time delay is indicated with the solid line, while the delayed one is indicated with the dashed line. The output signal from the M cell is the enhanced motion cue. 93
- 5.2 Schematic illustration of signal processing in the proposed DLMCE model. The pipeline of signal processing is shown in the left column with the lamina layers explained in further detail in the middle column. In the retina: firstly, the dark adaptation processing in the L layer is responsible for enhancing the dim-light image's intensity and contrast; then, the P layer calculates the luminance changes and implements spatio-temporal constraints for obtaining strong cues. In the lamina: the signal processing is the neural summation, including temporal summation in the E/I layer, spatio-temporal summation in the S layer and temporal summation in the M layer; specifically, the spatio-temporal summation involves the summation of lateral excitations and self-inhibition. Faint motion cues from dim-light input images are extracted and highlighted by the canonical neural computations in the retina and further enhanced by the neural summation processing in the lamina. 94

- 5.3 Illustration of the spatio-temporal constraint on luminance changes. (a) It displays the temporal constraint processing which exerts an adaptive threshold on changes in input luminance. It displays the temporal constraint processing which exerts an adaptive threshold on changes in input luminance. The baseline1 and baseline2 represent the adapted average intensity value at different times. The extracted luminance changes indicate that small fluctuations of input luminance are well inhibited. (b) It exhibits the spatial constraint processing where the normalization is implemented for the inputs. Outputs in the x axis at different instants in time t are constrained in a fixed range ($0 \sim 1$), which ensures stable outputs in spite of the overall luminance changes caused by different light levels. Note that the x coordinate represents the spatial position of image pixels. 97
- 5.4 Illustration of S layer's spatio-temporal summation mechanism. The excitatory (E) unit at the center position adds excitations from neighboring E units and the inhibition from the inhibitory (I) unit. Its output is then passed to the summation (S) unit in the same retinotopic position. The cell that transfers the signal with time delay is marked with dashed lines. The signals of delayed lateral excitations are indicated by the green arrows whereas the delayed self-inhibition signal is indicated by the red arrow. For clear illustration, only one S cell's summation processing is shown here. 98
- 5.5 (a)-(c): Various outputs during dark adaptation's processing with respect to the adaptation parameter. In each subplot, the horizontal axis denotes input intensity $I(x, y, t)$ while the vertical axis from left to right represent the output of Equ. (5.2), Equ. (5.3) and Equ. (5.4) respectively. In the left column, (a) $n = 0.4$, (b) $n = 0.5$, and (c) $n = 0.6$. Note that the value of I_m is manually set to 1, 50 and 150 individually representing different levels of background luminance. $L_o(x, y, t)$ is the final output for dark adaptation processing. 103
- 5.6 Sinusoidal grating images with different average intensity and/or RMS contrast as inputs, as well as outputs by using various image enhancement methods. (a) Input images. (b) CLAHE [19]. (c) LIME [20]. (d) Tanaka's [21]. (e) Proposed DLMCE's dark adaptation. 104
- 5.7 RMS contrasts of the input images (see Fig. 5.6(a)) are plotted against the RMS contrasts of the outputs with (a) CLAHE [19], (b) LIME [20], (c) Tanaka's [21], and (d) Proposed DLMCE's dark adaptation. 105
- 5.8 Four images of the same scene with different levels of luminosity and corresponding output images by using various image enhancement methods. (a) Dim-light images as inputs. (b) CLAHE [19]. (c) LIME [20]. (d) Tanaka's [21]. (e) Proposed DLMCE's dark adaptation. 106
- 5.9 Schematic illustration of the enhancement procedure of motion cues in the DLMCE model: a dim light image, as input, is fed into the L layer of the DLMCE model implementing dark adaptation processing; the changes in luminance are computed for extracting motion cues; these extracted motion cues are further enhanced by the subsequent spatio-temporal constraint and neural summation processing. 107

- 5.10 From top to bottom row: (a) Input images. They are sample frames from the nine dim-light image sequences, which contain a looming black disk under backgrounds over a range of average luminance ($13 \sim 63$) and RMS contrast ($0.07 \sim 0.35$). (b) Enhanced images. The dim-light images are firstly enhanced by the dark adaptation. (c) The enhanced motion cues. Motion cues are further extracted and enhanced by the spatio-temporal constraint and the neural summation. (d) Detected motion cues (moving edges). The enhanced motion cues in (c) are overlaid on the input images (a), which are visualized by red dots. Motion cues of the looming black disk from different dim-light levels of backgrounds can be effectively extracted by the proposed DLMCE model. 108
- 5.11 Schematic illustration of comparative experiments. From top to bottom: the raw input under normal illumination can be used to extract motion trajectory for obtaining ground truth; the raw input under dark/very dark illumination directly feeds into other motion cues detection methods for obtaining motion cues; the raw input is firstly enhanced by image enhancement methods (CLAHE/LIME/Tanaka's), and then feed into other motion cues detection methods for obtaining motion cues; the raw input feed into the proposed DLMCE model for obtaining motion cues. 109
- 5.12 Sample frames from synthetic datasets, including translating and looming objects under normal, dim and very dim illumination of backgrounds. Specifically, the degree of the background's darkness is controlled by the Matlab gamma intensity correction function, and the maximum output intensity is limited to the range of $[12, 102]$. From top to bottom rows: a translatory/looming colorful textured disk and a translatory/looming black one are shown in the sample frames. Note that the normal illumination dataset is used for extracting ground truth. Dim and very dim datasets as inputs feed into motion cues detection methods for conducting comparative experiments. 111
- 5.13 Results of motion cues detection from the first sequence of the very dim group (see Fig. 5.12), which show a detailed evaluation between the proposed DLMCE model and the comparative methods. 112
- 5.14 Comparison results of detected translating motion cues overlaid on the dim inputs between the proposed DLMCE model and other methods (**tested sequence one, a translatory black ball in a dim environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE. 113
- 5.15 Comparison results of detected translating motion cues overlaid on the very dim inputs between the proposed DLMCE model and other methods (**tested sequence two, a translatory black ball in a very dim environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE. 114

- 5.16 Comparison results of detected looming motion cues overlaid on the dim inputs between the proposed DLMCE model and other methods (**tested sequence three, a looming black ball in a dim indoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE. 114
- 5.17 Comparison results of detected looming motion cues overlaid on the very dim inputs between the proposed DLMCE model and other methods (**tested sequence four, a looming ball in a very dim indoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE. 115
- 5.18 Comparison results of detected translating motion cues overlaid on the dim inputs between the proposed DLMCE model and other methods (**tested sequence five, a translatory person in a dim outdoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE. 115
- 5.19 Comparison results of detected looming motion cues overlaid on the very dim inputs between the proposed DLMCE model and other methods (**tested sequence six, a looming person in a very dim outdoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE. 116
- 6.1 The proposed bio-inspired dim-light collision detection model is composed of five layers (DLMCE, P_M , E_M , I_M , S_M), and the LGMD1 cell and the feedforward inhibition (FFI) cell. Specifically, it combines the DLMCE model introduced in Chapter 5 with the classical four-layered LGMD1 model [5], [22] together where the outputs of DLMCE feed the first layer of the LGMD1. 121
- 6.2 Sample frames from the four tested image sequences in the dim-light dataset. From the first column to the fourth column are frame 50, frame 100, frame 150 and frame 200 separately. From top row to bottom row are sequence one to sequence four: (a) sequence one, a looming white ball; (b) sequence two, a looming black ball; (c) sequence three, an approaching woman wearing a light T-shirt; (d) sequence four, an approaching man wearing a dark T-shirt. It is worthy to note that sample frames in (a), (b) and (c), (d) are captured in indoor and outdoor environments individually. 130

- 6.3 Experimental results of the proposed model for sequence one (subfigure (a)) and sequence two (subfigure (b)). In each subfigure, it includes the dark image (input), enhanced intensity image (output of L sublayer in DLMCE layer), enhanced motion cues (output of M sublayer in DLMCE layer) and response of LGMD1 neuron (NMP curve). Note that each exhibited image in each subfigure is the result of frame 200. 131
- 6.4 Experimental results of the proposed model for sequence one (subfigure (a)) and sequence two (subfigure (b)). In each subfigure, it includes dark image (input), enhanced intensity image (output of L sublayer in DLMCE layer), enhanced motion cues (output of M sublayer in DLMCE layer) and response of LGMD1 neuron (NMP curve). Note that each exhibited image in each subfigure is the result of frame 200. 132
- 6.5 The comparison of LGMD1 neural response among Rind's model [5], Yue's model [7] and the proposed model for sequence one and two. The output image of the last layer in the model with respect to the alarm frame is also presented. Note that the alarm frame was produced by the proposed model. **The detected collision alarm frame in sequence one (a): Rind's model (Frame 199), Yue's model (Frame 192) and Proposed model (Frame 162). The detected collision alarm frame in sequence two (b): Rind's model (No alarm frame), Yue's model (Frame 200) and Proposed model (Frame 169).** 134
- 6.6 The comparison of LGMD1 neural response among Rind's model [5], Yue's model [7] and the proposed model for sequence three and four. The output image of the last layer in the model with respect to the alarm frame is also presented. Note that the alarm frame was produced by the proposed model. **The detected collision alarm frame in sequence three (a): Rind's model (No alarm frame), Yue's model (No alarm frame) and Proposed model (Frame 122). The detected collision alarm frame in sequence four (b): Rind's model (No alarm frame), Yue's model (No alarm frame) and Proposed model (Frame 163).** 135

List of Abbreviations

LGMD	Lobula giant movement detector
LMC	Lobula monopolar cell

Chapter 1

Introduction

1.1 Motivation

Collision detection is critical for autonomous vehicles or robots to serve human society safely. However, it is challenging to correctly recognize looming objects in natural scenes, especially in dim light environments. Insects in the biological world have evolved exquisite vision systems to detect approaching predators or obstacles within complex visual scenes, thus enhancing their chances of survival [23]–[25]. For example, the lobula giant movement detector (LGMD) is an identified neuron in the locust visual system, which can respond preferentially to looming objects and initiates escape behaviors [26]. The LGMD neuron is shown in Fig. 1.1(a). In low light conditions, the captured image has low luminance and low contrast, as shown in Fig. 1.1(b). The motion is difficult to detect for an artificial visual system since the motion cues are very weak under such illumination conditions. However, many insects are still active at night and can detect faint movements like negotiating dimly illuminated obstacles during flight [27], [28]. Hence, insects’ compacted brains and specialized neural mechanisms can provide an abundant source of inspiration for developing specific functionalities such as navigation, tracking, and collision detection in autonomous robots and vehicles. Specifically, it is essential to explore the LGMD’s looming sensitivity and insects’ nocturnal vision for achieving robust collision detection.

The looming-sensitive LGMD1 (used to be named LGMD) neuron can be excited by an approaching, light or dark object, and trigger escape or avoidance behaviors [29].

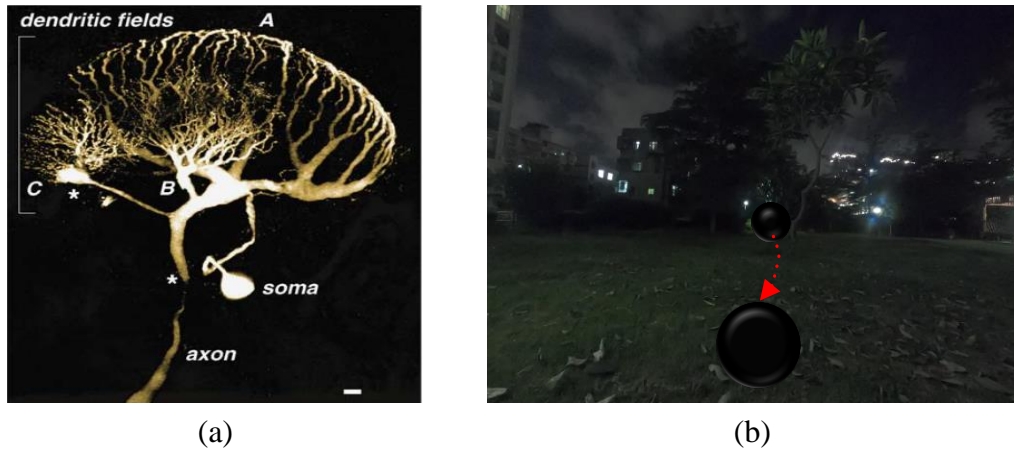


Figure 1.1: (a) The LGMD neuron's dendritic tree, which consist of three distinct subfields (A–C). This figure is adapted from [1]. (b) A looming black ball under dim light environment.

Many LGMD1-based artificial vision systems have been developed to detect objects approaching on a collision course [7], [8], [30], [31]. However, existing models cannot distinguish looming objects from near and fast translating objects. For example, a fast translationally moving car passing in front of the video camera evokes false LGMD1 spikes and produce unwanted collision alarms [32]. A pedestrian walking on a pavement may also trigger a false alarm due to rapid translational movement [6]. How the presynaptic neural processing between inhibitions and excitations should contribute to shaping the selective responses of the LGMD1 model to looming objects instead of fast translating ones, remain unclear. Neural information processing mechanisms, such as lateral inhibition [5], spiking frequency adaptation (SFA) [33], separated processing of ON/OFF signals [8], and synaptic plasticity mechanisms [26], have been applied in previous models. However, they have not been able to deal with the challenges presented by translating stimuli, such as fast translationally moving objects traveling at loom-like speeds or moving in a divergent mode. In biology, the neural competition between ON and OFF channels can explain response features arising from opponent inputs [34]. Also, neurons compete with each other make the winner ones tuned to the certain pattern of inputs but keep the others from becoming selective to that same pattern [35]. Therefore, it is possible to propose a new LGMD1 model to enhance collision selectivity by comparing ON and OFF responses.

In addition to the challenges in shaping the directional selectivity for looming mo-

tion, there are further challenges in collision detection under low light conditions. Due to the scarcity of photons, the captured low-light images with low intensity and contrast usually accompany extremely low signal-to-noise ratios (SNRs). Movements of objects are buried in a poor lighting environment, and motion cues are correspondingly very weak. Motion cues, as visual stimuli, can trigger directionally selective neurons that respond to specific movement patterns like the lobula giant movement detectors (LGMDs) [29], [36]. The extraction of motion cues is a fundamental processing step in any visual system in nature [37] or for autonomous systems like robots or vehicles. At high luminance levels, it is easy for artificial vision systems to recognize motion cues as the SNRs of the acquired visual stimuli are high [32], [38]–[41]. However, motion cues are faint and hard to obtain under low light conditions, which makes the detection of motion very difficult for autonomous vehicles or robots. As a result, the enhancement of motion cues is sorely needed.

Since nocturnal animals show remarkable visual abilities in detecting faint movements [27], [28], specialized visual processing strategies within their brain can provide inspiration for exploring the functionality of enhancing motion cues. For instance, researchers have widely investigated the neural summation strategy and have adopted it to develop models for low-light image enhancement [10], [11], [42], [43]. More concretely, these methods mainly apply the idea of neural summation for noise reduction as the intensity transformation processing greatly amplifies the noise. The summation strategy has verified its effectiveness in reducing the inevitable noise, however, the computation of the summation kernel is computationally expensive. To enhance motion cues, it requires the luminance distortion as less as possible as the motion cues are related to changes in luminance. It implies that the processing for raising the dark pixel's intensity should preserve its naturalness. The dark adapted photoreceptors [12] and the canonical neural computations [15] can explain light adaptation in the retina, which is applicable for exploring the adaptive intensity transformation algorithm. Moreover, the neural summation strategy in temporal and spatial domains can be applied to enhance motion cues. In brief, the biological solution for motion cues enhancement makes it possible to develop an effective and efficient visual system for

collision detection in dim light environments.

To enhance the collision selectivity of the existing LGMD1 models, this thesis develops a new LGMD1 model with neural competition between ON and OFF channels that can effectively distinguish looming stimuli from translatory ones. To enhance the low-light image and preserve the naturalness, this thesis proposes a dark adaptation framework that can adaptively raise the image intensity. To reliably detect the moving objects in low light conditions, this thesis develops a DLMCE model to enhance dim-light motion cues by adopting canonical neural computation and neural summation strategies on the basis of the proposed dark adaptation framework. To detect potential collisions in low light conditions, this thesis proposes a bio-inspired dim-light collision detection model that combines the DLMCE model and the classical four-layered LGMD1 model for detecting looming objects.

1.2 Organization of the Thesis

This thesis consists of seven chapters. Chapter 1 gives the motivation and organization of the thesis. Chapter 2 summarizes the related work on collision detection and motion cues enhancement methods. The main contributions of this thesis are contained in Chapter 3, 4, 5 and 6, which present the designed LGMD1, framework of low light image enhancement, the DLMCE, and the bio-inspired low-light collision detection model respectively. In each of these four chapters, we will discuss the design motivation, describe the proposed model or method, and demonstrate the experimental results. Chapter 7 finally summarizes the main contributions of this thesis and outlooks the possible future work. The details in each chapter are summarized as follows.

Chapter 2 mainly reviews the related work on collision detection and motion cues enhancement. It first introduces biologically inspired collision detection methods. Then the traditional collision detection methods are summarized. Furthermore, it reviews the related work on motion cues enhancement for collision detection in dim-light environments, including neural mechanisms in nocturnal insects' visual systems, low-light image enhancement methods, and motion cues detection methods.

Chapter 3 develops a new LGMD1 neural network model for looming cue detection

against translating motion. It first introduces the background and indicates shortages of the existing LGMD1 models. Then the newly developed neural network model is formulated. Finally, the experimental results are presented and discussed.

Chapter 4 proposes a dark adaptation framework for enhancing dim-light images. It first introduces the background and overview of this study. Then the proposed dim-light image enhancement method is described. Finally, the experimental results are demonstrated and discussed.

Chapter 5 develops a dim-light motion cues enhancement (DLMCE) model for enhancing faint motion cues. It first introduces the background and indicates the existed problems. Following that, the structure and formulation of the proposed DLMCE model are detailed. After that, we present the experimental results and discussions.

Chapter 6 proposes a bio-inspired collision detection model for detecting looming objects in dim light environments. The formulation of the model is first introduced, and its components are elaborated. Then, the experimental results are displayed, involving the results comparison of collision detection between the proposed model and the existing LGMD1 models.

Chapter 7 summarizes the main research contributions of this thesis and discusses the potential work that can be further carried out in the future.

Chapter 2

Related Work

For an artificial collision detection system, the ability to detect looming cues is vital. Insects' visual systems have provided a rich source of inspiration in developing collision detectors since they are capable of perceiving colliding objects in complex dynamic scenes by their directionally selective neurons involving looming-sensitive neurons. In particular, the wide-field looming-sensitive neuron LGMD1 has been widely studied and modeled. For collision detection in low light conditions, motion cues are so weak that an effective enhancement for the faint motion cues is needed. Nocturnal animals' specific neural mechanisms are reasons for seeing well in dim light that can be applied to enhance motion cues.

This chapter mainly reviews the related work on collision detection approaches and motion cues enhancement methods. It begins with an introduction of biologically inspired collision detection methods in Section 2.1, involving directionally selective neurons and their computational models for collision detection with LGMD1 based models detailed discussed. Then, traditional collision detection methods are also summarized and discussed in Section 2.2. Following that, Section 2.3 reviews relevant neural mechanisms in nocturnal animals' visual systems. Finally, Section 2.4 and 2.5 summarize low-light image enhancement and motion cues detection methods respectively, and discuss the limitations of low-light image enhancement approaches in motion detection.

2.1 Biologically Inspired Collision Detection Methods

The ability to detect motion is ubiquitous amongst most animals' visual systems. Motion-sensitive neurons discovered have been extensively studied [29], [37], [44]–[47]. Various bio-inspired motion detectors have been proposed correspondingly, such as collision detectors [5], [7], [39]–[41], [48]. The biologically inspired collision detectors extract looming information by mimicking directionally selective neurons, including looming-sensitive neurons. In this section, we first review directionally selective neurons and their applications for collision detection. Then, models based on the widely investigated looming detector LGMD1 are described and discussed in detail.

2.1.1 Directionally Selective Neurons

2.1.1.1 Wide-field-motion-sensitive Neurons

Lobula plate tangential cells (LPTCs) found in the lobula plate of the fly visual system are sensitive to visual motion in a directionally selective way [37], [47], [49]. These LPTCs are divided into two systems [50], [51]: one system consists of the horizontal direction-sensitive (HS) cells, that depolarise rigorously in response to horizontally moving stimulus in the preferred direction while hyperpolarizing when stimulated by motions along the opposite direction; another system consists of the vertical direction-sensitive (VS) cells, and the response properties are similar to the former system.

The motion detection mechanism of HS cells and VS cells have been extensively investigated [48], [52], [53]. And the elementary motion detector (EMD) is a well-known computational model for motion detection based on the motion mechanism of HS cells and VS cells. Specifically, the correlation-type EMD (so-called Reichardt model) correlates the adjacent delayed and non-delayed signals. Fig. 2.1 illustrates an EMD process in four steps, which explains how an EMD can produce a strong directionally selective signal by the activity of neighboring photoreceptors. Moreover, Fig. 2.2 exhibits four classical elementary motion detectors (EMDs) that devote to disclose the motion detection mechanisms within the fly visual system.

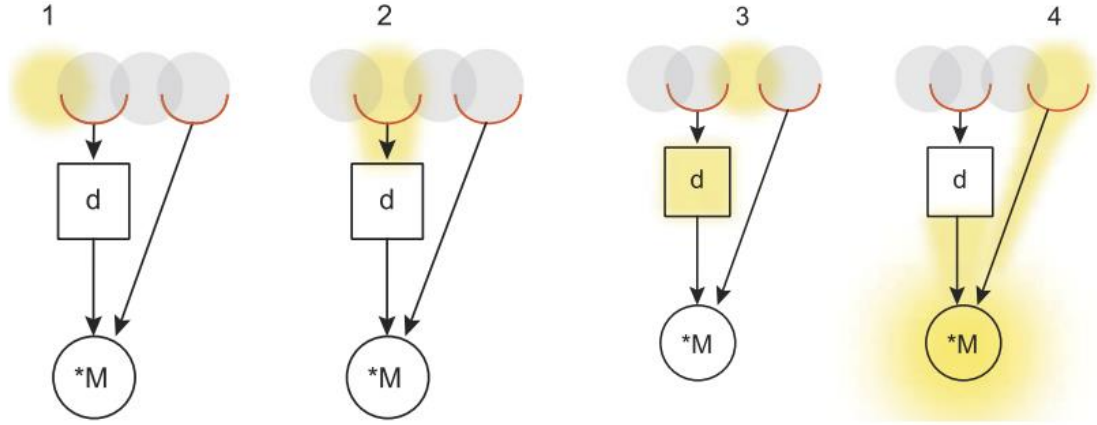


Figure 2.1: Schematic illustration of the EMD process in four steps. The figure is adapted from [2].

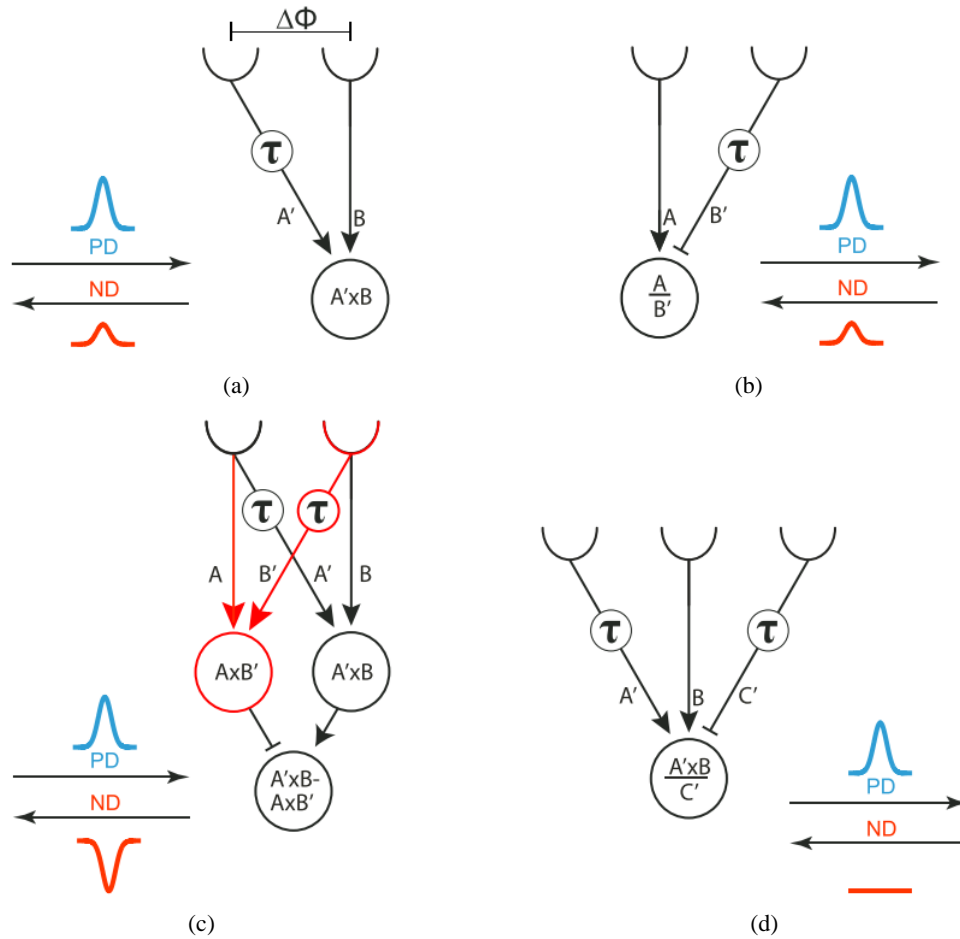


Figure 2.2: Schematic illustration of four classical EMD models. (a) Hassenstein-Reichardt-(half)-detector: preferred-direction (PD) enhancement. (b) Barlow-Levick-detector: null-direction (ND) suppression. (c) (full) Hassenstein-Reichardt-detector. (d) HR/BL-detector: PD-enhancement + ND-suppression. The figure is adapted from [3].

Since the EMDs can measure expanding optic flow, they are used to develop collision detection systems [54]–[56]. Furthermore, EMDs are used to analyze directional excitation patterns as an assistant method in an LGMD-based collision detection system [57]. However, the EMD-based collision detection methods depend on the reliable computation of optical flow, and the coding of optical flow is always ambiguous [58]. The reason is that the two parameters in the EMD — the spacing between a pair of detectors and the temporal delays for both HS and VS detectors — can significantly influence the detection of motion direction and intensity [59].

2.1.1.2 Looming-Sensitive neurons

There are identified looming-sensitive neurons in animals (locust, pigeon and mouse) that respond selectively to looming motion stimuli [36], [44], [60], [61]. These looming-sensitive neurons can protect the animal from collision by eliciting protective motor responses, which is essential for the animal’s survival. Researchers have proposed many computational models for modeling these looming sensitive neurons, especially the LGMD1 based neural network models [5], [7]–[9], [62]–[65]. More importantly, extensive studies about the LGMD1 neuron make the LGMD1 a good choice in developing the bio-inspired collision detection system [1], [24], [45], [66], [67]. A detailed review of LGMD1 is given below.

The LGMD1 is one of a group of giant movement-detecting neurones, which have fan-shaped arbors in the lobula of the locust optic lobe and respond directionally to movements of objects in depth [29]. The LGMD1 and its target neuron, the descending contralateral movement detector (DCMD), make up a visual processing pathway that responds preferentially to approaching objects [68], [69]. Fig. 2.3 illustrates the locust visual system where the LGMD1 synapses onto the DCMD and spikes in the DCMD follow those in the LGMD1 one for one [70]. Since the LGMD1 neuron plays a dominant role for locusts’ flying with low collision rates, studies mainly focus on exploring the LGMD1 [4], [71]. Next, we will discuss the existing LGMD1-based models.

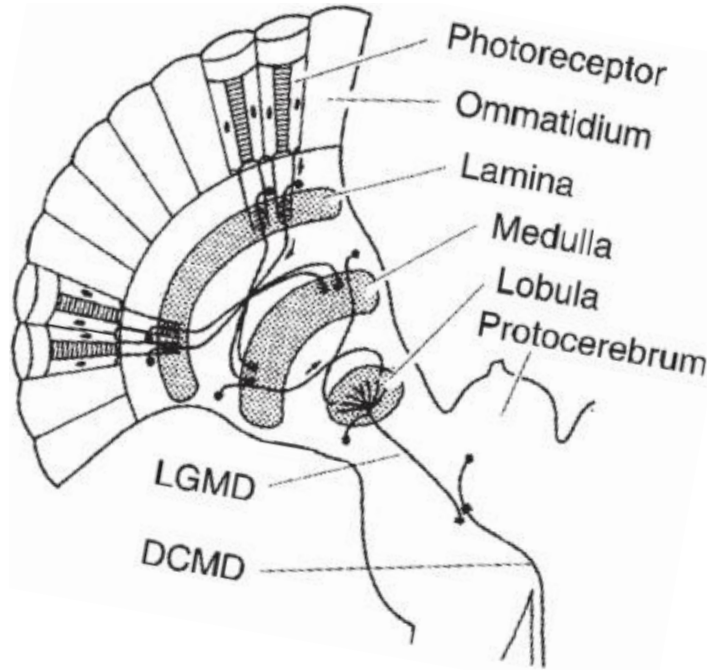


Figure 2.3: Schematic illustration of the locust visual system: Presynaptic neuropile layers of the LGMD1 neuron and its postsynaptic target neuron, the descending contralateral movement detector (DCMD). The figure is adapted from [4].

2.1.2 LGMD1 based Neural Network Models

Over several decades, there were two representative ways of modeling the LGMD1 neuron. Rind and Bramwell [5] proposed one way that models the LGMD1 based on the input organization using lateral inhibition within the presynaptic network to shape the collision selectivity. Later, many LGMD1 model variants and applications based on this way occurred [7], [22], [31], [32], [38], [72]. While Gabbiani et al. [73] proposed another way that shapes the LGMD1's collision selectivity by the non-linear interactions between the excitations and inhibitions in a mathematical perspective where the feedforward excitation/inhibition is related to the angular speed/size of looming objects within the visual field [1], [45]. Other LGMD1 model variants were built based on this way [74]–[76]. For the LGMD1 modeling, there should be many possible ways to build the collision selectivity by elucidating the biophysical mechanisms underlying LGMD1's responses to looming objects from different viewpoints [73]. In this thesis, we try to enhance the existing LGMD1's looming sensitivity from the perspective of input organization. Therefore, this type of LGMD1 model variants are highlighted, and the detailed discussions of these models are presented below.

The first LGMD1 neural network model proposed by Rind and Bramwell [5] is

shown in Fig. 2.4. The LGMD1 model utilizes the lateral inhibition to filter out non-looming visual stimuli, such as receding and translating stimuli. Specifically, selective responses to looming objects is achieved by altering the strength of lateral synaptic connections and the timing of inhibition signals. The lateral inhibition can inhibit responses elicited by receding or translating stimuli using this way. However, the effect of the lateral inhibition is affected by the moving object's size and velocity. For example, large and rapid translational movements can also trigger strong responses since the most translational excitations cannot be effectively suppressed. It indicates that this LGMD1 model may lead to produce false collision alarms under certain circumstances.

To address this issue, the way of adjusting parameters [32] or integrating networks that were sensitive to translational movements [6], [57] was adopted to reduce false collision alerts. The LGMD1 based neural network combined with the asymmetric lateral inhibition based translating sensitive neural network (TSNN) is shown in Fig. 2.5. Nevertheless, the problem of distinguishing between looming objects on a collision course and translationally moving objects remained. Moreover, the combined model demands more computational resources and does not reflect the intrinsic collision-sensitive properties of the biological LGMD1 neuron.

There are other improvements for the first LGMD1 model. Yue and Rind [7] developed new mechanisms in the LGMD1 based visual neural network, as illustrated in Fig. 2.6. It adds a G layer after the S layer, which can enhance the ability to extract looming features from complex and dynamic environments. Still, the collision selectivity for the LGMD1 based model is an open challenge. Fu et al. [8] proposed a new LGMD1 model, as illustrated in Fig. 2.7. The model utilizes separate ON/OFF pathways to process the visual stimuli and combined with the spike-frequency adaptation (SFA) mechanism [33], [77], [78] to shape the collision selectivity. However, its biophysical SFA mechanism is modeled by computing the derivative of a neural response as well as two time-dependent coefficients. This makes the model sensitive to changes in an object's velocity. Responses to translating objects at a constant speed can be effectively inhibited. But those with changing velocity may elicit strong responses. It indicates that there is an additional inhibitory mechanism for differentiating looming

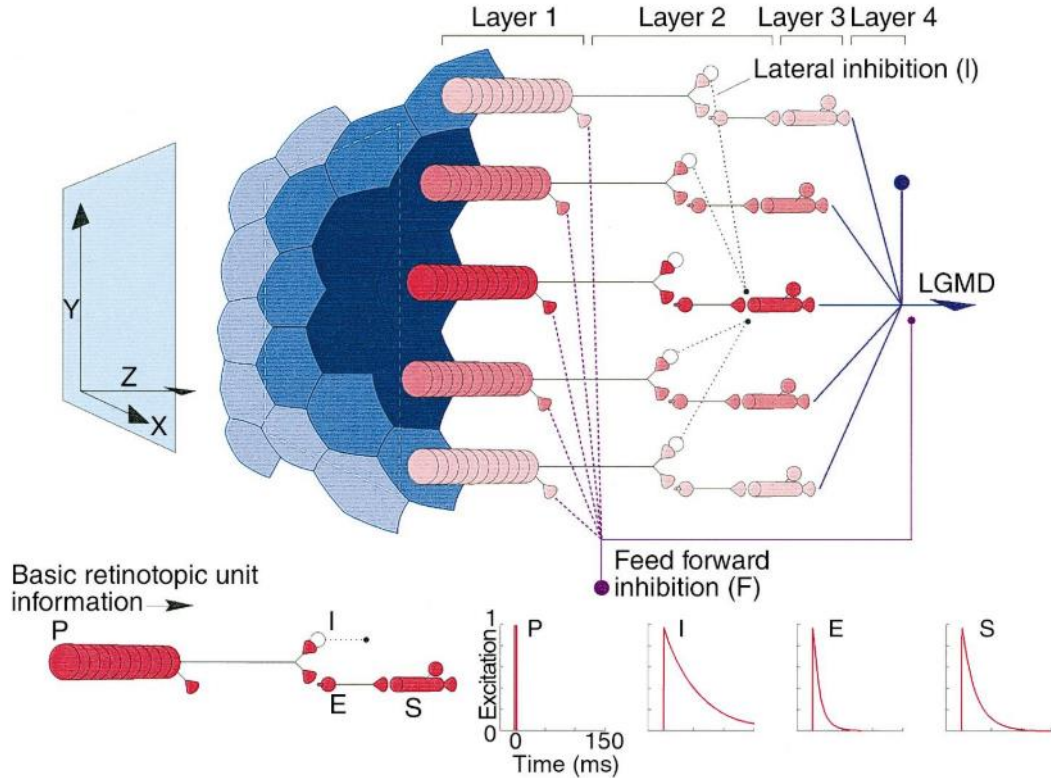


Figure 2.4: Schematic of LGMD1 visual neural network proposed by Rind and Bramwell [5]. The model consists of four layers, including a LGMD1 cell and a FFI cell. The input organization of the basic retinotopic unit of the network is shown in red and labeled. The outputs of these units were summed by the LGMD unit (blue). In layer 1, images were mapped onto the photoreceptors. Each photoreceptor (P unit, layer 1) in the model responded with a brief excitation to a change in level of illumination (graph P). In layer 2, this excitation was passed to three units: E, I and F. When excited, E passed excitation to a layer 3, S unit in the same retinotopic position. I passed delayed inhibition laterally to the nearest S units in the layer 3. F fed inhibition forward and by-passed layer 3. The E and I inputs were summed linearly by each S unit in layer 3 until a threshold level of excitation was reached and a spike was produced. Layer 4 of the model consisted of a single LGMD unit that summed excitation from all active S units and inhibition from the F unit. In each layer proximity to the central retinotopic unit is indicated by the shade of red (dark red is the nearest). The figure is adapted from [4].

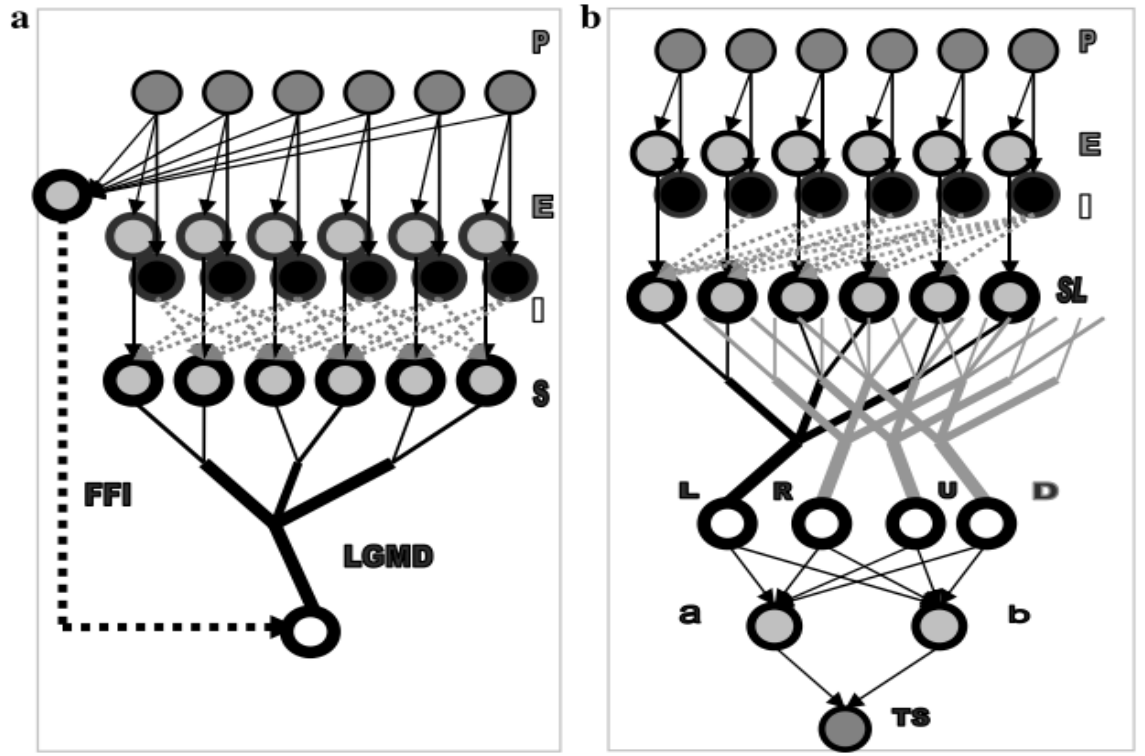


Figure 2.5: The schematic illustration of the LGMD1 based neural network and the asymmetric lateral inhibition based TSNN for colliding objects detection. (a) The LGMD1 based neural network. (b) The asymmetric lateral inhibition based TSNN. The figure is adapted from [6].

from translation [78]. For the models with separate ON/OFF pathways, responses to looming and translational motion should be quite different. As a result, we proposed new LGMD1 models based on the neural competition between the ON and OFF channels. Note that initial version of the neural competition based LGMD1 model is shown in Fig. 2.8. The LGMD1 model with neural competition mechanism can effectively inhibit responses elicited by translational movements, which enhances the collision selectivity.

2.2 Traditional Collision Detection Methods

For collision detection, accurate estimation of motion from a sequence of images is the key. [79] has given an intuitive illustration for motion that a moving pattern of light falls upon the retina when the eye is in motion relative to the visible environment, and the motion pattern in 2D video imagery contains distance information about objects in a 3D environment [80]. In other words, there is apparent movement in a sequence

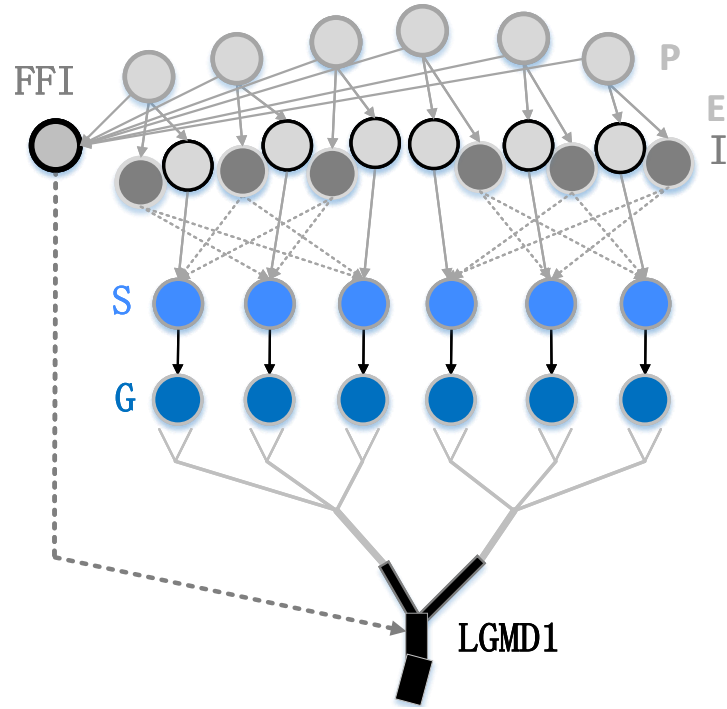


Figure 2.6: The LGMD1 model proposed by Yue and Rind [7]. There are five groups of cells and two single cells: Photoreceptor cells (P); excitatory and inhibitory cells (E and I); summing cells (S); grouping cells (G); the LGMD cell; and the feed forward inhibition cell (FFI). The input of the P cells is the luminance change. Lateral inhibition is indicated with dotted lines and has one frame delay. Excitation is indicated with black lines and has no delay. The FFI also has one frame delay. The input to FFI is the luminance changes from photoreceptor cells.

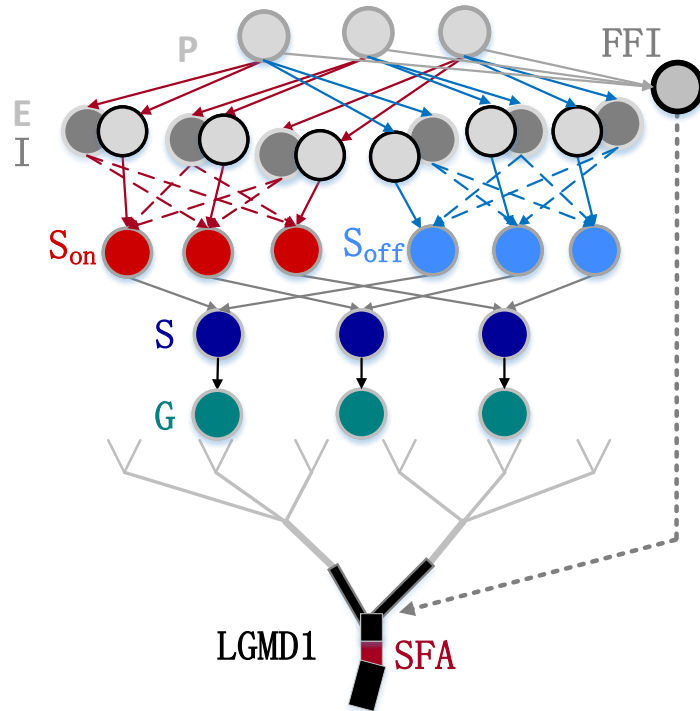


Figure 2.7: The LGMD1 model proposed by Fu et al. [8]. The proposed LGMD1 model processes signals in the ON and OFF pathways without bias, whilst a new SFA mechanism is modeled. The dashed lines indicate transmissions of delayed neural signals.

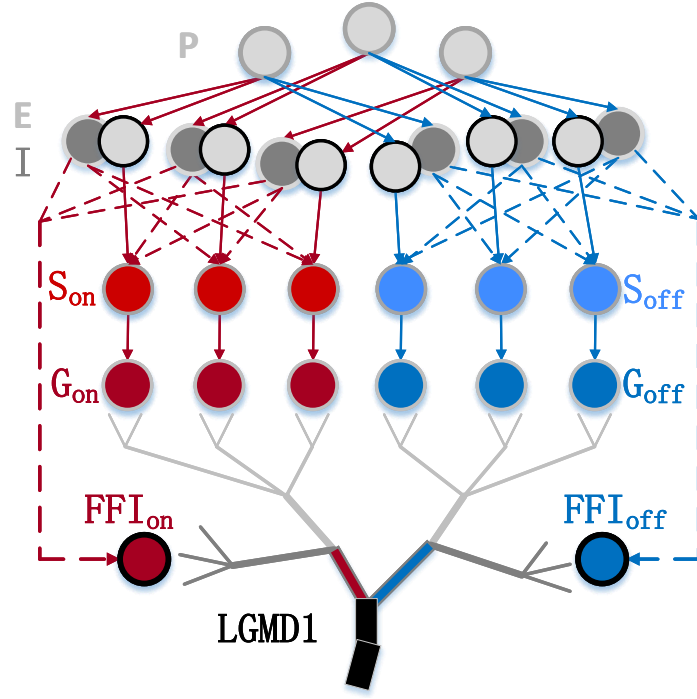


Figure 2.8: The neural competition based LGMD1 model. The model is composed of five layers (P, E, I, S, G) and three cells (FFI_{on}, FFI_{off}, LGMD1); signals are split into ON (red-arrows) and OFF (blue-arrows) pathways each with four layers (E, I, S, G); excitatory signals make a neural competition in LGMD1 neuron. Note that all transmitted signals with time delays are indicated with dashed lines. The figure is adapted from [9]

of images while the relative motion occurs between objects in a scene and a camera [81], which can be represented by a discrete set of features or brightness patterns. Traditional methods for collision detection aim to obtain crucial parameters like the time-to-collision by analyzing the motion patterns.

Specifically, optical-flow based and feature based are two highlighted collision detection methods. The former estimates the time-to-collision or time-to-contact (TTC) from the optical flow to detect obstacles. The latter calculates the expansion ratio of relative feature sizes and/or areas to recognize approaching objects. The performance of the two types of traditional methods strongly relies on the accurate estimation of optical flow and extraction of feature points. However, it is rather challenging to estimate the optical flow and the feature points under certain conditions like textureless. A detailed review of the optical-flow based and the feature-based is given below.

2.2.1 Optical-flow based methods

The time to contact (TTC) can guide navigation to avoid potential collisions by predicting the time to reach an object in front of them. The optical-flow based methods calculate the TTC from specific information provided by the optical flow to achieve collision detection. Next, we first introduce the classical optical flow methods and then review the optical-flow based collision detection methods.

2.2.1.1 Optical Flow

Optical flow can provide important information about the spatial arrangement of the objects viewed and the rate of change of this arrangement [82]. Optical flow was defined as the distribution of apparent velocities of movement of brightness in an image [83]. The goal of estimating optical flow is to compute an approximation to the motion field from time-varying image intensity [84]. Numerous methods have been proposed to estimate optical flow, such as differential methods, region-based matching, energy-based and phase-based techniques, etc.. Differential methods are the most widely used techniques [85], which can be classified into global methods such as the Horn/Schunck (HS) approach [83] and local methods such as the Lucas-Kanada (LK) technique [86], [87]. The HS and LK algorithms are described as follow.

HS algorithm: First, the change in image brightness at a point is considered related to the motion of the brightness pattern. Let the brightness of a pixel point in the image plane at time t is denoted by $E(x, y, t)$. Then, the brightness of the pixel point in the pattern is assumed to be constant when the pattern moves. The brightness constancy constraint is given by,

$$\frac{dE(x, y, t)}{dt} = 0 \quad (2.1)$$

Using the chain rule for differentiation, we see that,

$$\frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \quad (2.2)$$

Let $u = \frac{dx}{dt}$, and $v = \frac{dy}{dt}$, then there is a single linear equation in the two unknowns u

and v ,

$$E_x u + E_y v + E_t = 0 \quad (2.3)$$

where E_x , E_y and E_t denote the partial derivatives of image brightness with respect to x , y and t respectively. The brightness gradient E_x and E_y are components of the optical flow, whilst u and v are components of the flow velocity. Note that the flow velocity cannot be computed locally without additional constraints. For example, the smoothness is adopted as an additional constraint as discontinuities in flow can be expected where one object occludes another.

LK algorithm: Lucas and Kanade proposed a more general measure of image difference. They assumed that the unknown optic flow vector (u, v) is constant within the neighborhood of the pixel point. In this case it is possible to determine the two *constants* u and v at some location (x, y, t) from a weighted least square fit by minimizing the function [88]

$$E_{LK}(u, v) := K_\rho * ((E_x u + E_y v + E_t)^2) \quad (2.4)$$

where ρ indicates the size of the neighborhood. A minimum (u, v) of E_{LK} satisfies $\partial u E_{LK} = 0$ and $\partial v E_{LK} = 0$. Since the size of the ρ is able to affect the accuracy and robustness, a pyramidal implementation of the LK algorithm was proposed [89].

2.2.1.2 OF-E Method for Collision Detection

The OF-E method [90] estimates the time-to-contact (TTC) from the optical flow (OF) that contains the information about the direction of the point of observation called the focus of expansion (FOE). The computation of TTC is given by

$$TTC = \frac{\Delta_i}{|\vec{V}_t|} \quad (2.5)$$

where Δ_i is the distance of the considered point (x_i, y_i) on the image plane from the focus of expansion (FOE), V_t indicates the optical flow that relates to the image velocity. Note that the OF-E method applies a standard HS algorithm [83] to compute the optical flow. The calculation of FOE is based on the principle that flow vectors are

oriented in specific directions relative to the FOE [90].

2.2.1.3 OF-D Method for Collision Detection

The OF-D method [91] estimates the time-to-contact (TTC) by the divergence of image flow. Divergence of the flow field is computed by the sum of image flow derivatives in two perpendicular directions. The flow divergence can give a measure of scene structure for a moving observer, which is inversely related to TTC [92]. Therefore, the TTC can be directly estimated from the flow divergence. That is,

$$TTC = \frac{Z}{V_Z} = \frac{2}{D} \quad (2.6)$$

where Z indicates the distance between the observer and the looming object, V_Z is the velocity of the looming object, D is the divergence of the image flow.

For computing the divergence, the Lucas-Kanade tracker [89] first tracks corners that are detected by the FAST algorithm [93], [94]. Then, the divergence is estimated by the following equation,

$$D = \frac{1}{n} \cdot \frac{1}{\Delta t} \sum_{i=1}^n \left[\frac{d_{(t-\Delta t),i} - d_{t,i}}{d_{(t-\Delta t),i}} \right] \quad (2.7)$$

where n is the total number of tracked corners, Δt is the time interval between two consecutive images. $d_{t,i}$ indicates the image distance between every two corners at one image, and $d_{(t-\Delta t),i}$ denotes the corner distance at the next image. The expansion or contraction of the flow is measured by further computing the ratio of $d_{(t-\Delta t),i} - d_{t,i}$ to $d_{(t-\Delta t),i}$. Finally, the value of the TTC is obtained by the calculation of the divergence (see Eq. 2.6).

2.2.2 Feature based methods

In order to detect the frontal obstacle, the relative size cue is utilized to predict the time to collision (TTC) since it is proportional to the response to the expansion of an obstacle. The feature-based methods aim to develop a relative size detector to detect and avoid imminent collisions. First, SIFT/SURF feature match is implemented. Then,

the relative obstacle size or area is compared. Finally, the position of the obstacle in the image is estimated for performing the avoidance maneuver. In the following, the SIFT/SURF detectors are first described, and the feature-based methods for collision detection are then reviewed [95], [96] .

2.2.2.1 SIFT/SURF Detectors

For a feature-based approach, a feature is required to be robust with respect to small visual changes (illumination, rotation, pose, and scale) [97]. The Scale Invariant Feature Transform (SIFT) detector [98] or a scale- and rotation-invariant detector and descriptor, coined SURF (Speeded-Up Robust Features) detector [99], are widely used feature descriptors. The SIFT and SURF feature descriptors satisfy the requirements for a feature-based method. The algorithm for SIFT involves the computation of gradient histograms, which divides the image region into 16 sub-regions (4×4) and computes a histogram of local orientation gradients around the interest point in each sub-region [98]. As a result, the SIFT descriptor contains 128 values (8 directional values for 16 sub-regions). The algorithm for SURF employs integral images and efficient construction of scale-space to generate keypoints and descriptors, which approximates SIFT but is much quicker to compute than SIFT [100].

2.2.2.2 FD-S Method for Collision Detection

The FD-S method [95] utilizes SURF feature matches combined with template matching to compare relative obstacle sizes with different image spacing. The SURF keypoints in consecutive images are matched, and the size of the region around the keypoint is compared. Note that only the keypoint's size becomes larger will be kept. In addition, the expansion ratio of SURF keypoint needs further confirmation by using the template matching. Finally, a group of expanding SURF keypoints are obtained for recognizing if there is an obstacle.

2.2.2.3 FD-A Method for Collision Detection

The FD-A [96] method estimates the change in the area size of an approaching object to detect a potential collision. It adopted the SIFT detector to identify and localize feature points. Other than the size comparison of matched points like the FD-S method, the FD-A method also compared the size of the convex hull (area) constructed by the detected feature points. When the size ratio exceeds a certain empirical value, it indicates that there is an obstacle detected.

2.3 Neural Mechanisms in Nocturnal Vision

Nocturnal insects have evolved a remarkable capacity to visually navigate at night due to their highly sensitive compound eyes and specialized visual processing strategies in the brain [28], [101]. The visual neural systems of nocturnal animals with well evolved mechanisms allow them to see objects in dim light environments and perceive faint motion cues. These mechanisms are mainly involving photoreceptors adapted to dark conditions for amplifying visual signals [12], canonical neural computations for adjusting the sensitivity of neurons through normalization [15] as well as the high-level neural summation for strengthening the enhanced signals [10]. The biological evidence and modeling works have indicated that these neural mechanisms contribute a lot to nocturnal vision, for example, some neural computation strategies have been used to develop specific functionalities in artificial vision systems, including modeling the adaptation of photoreceptors [102] and the noise-reduction of video [11]. In this section, we review the related work of neural mechanisms in the following order - photoreceptors' dark adaptation first, then the canonical neural computations, and finally the neural summation.

2.3.1 Photoreceptors' Dark Adaptation

Nocturnal insects can see well the same as their diurnal relatives in the complex visual world due to their highly sensitive eye designs and remarkable neural adaptations [101], [103]. The biological evidence from both anatomical and physiological

characteristics of the eye demonstrated that the reliability of vision starts with photoreceptors [104]. Next, the designs of compound eyes and dark-adapted photoreceptors are introduced.

2.3.1.1 Designs of Compound Eyes

The compound eye consists of repetitive visual units (ommatidia), each of which contains a dioptric apparatus that focuses light onto the layer of photoreceptors [105]. Designs of two types of compound eyes are shown in Fig. 2.9, which can explain the compound eyes' optical adaptation. As can be seen from Fig. 2.9 (a), the number of

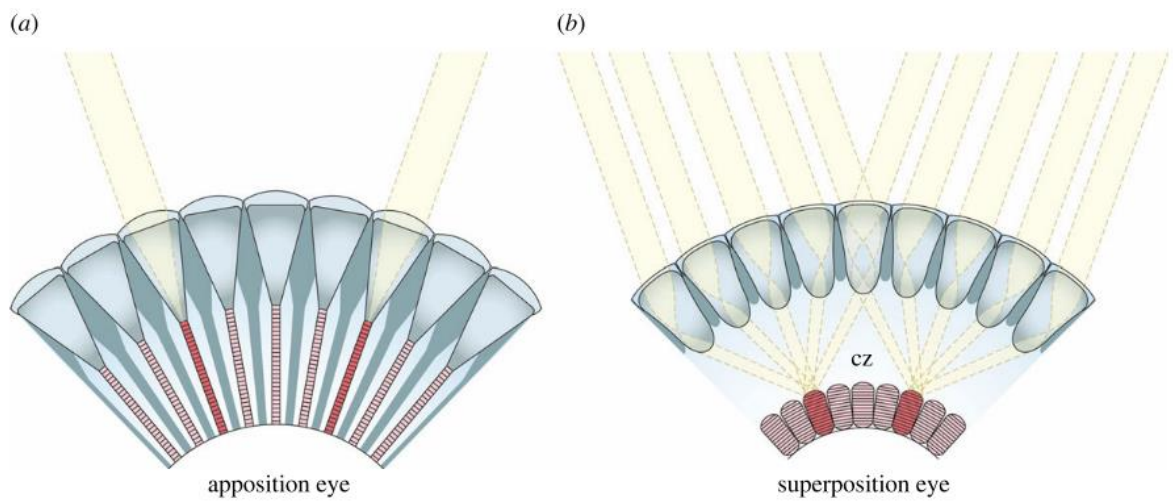


Figure 2.9: (a) Apposition compound eyes. There are nine ommatidia here. The photoreceptors receive light exclusively from the single facet lens of their own ommatidium. (b) Superposition compound eyes. The lenses and photoreceptors are separated by a wide optically clear region known as the clear zone (cz). The lenses allow light from many hundreds of facet lenses to be focused onto single photoreceptors in the retina, thus dramatically increasing light capture. The figure is adapted from [10].

captured photons from the apposition compound eyes is limited due to the small size of the individual facet lens. Each ommatidium, therefore, views a 'pixel' of the overall image. Most day-active insects possess the apposition compound eyes [106]. The superposition compound eyes are shown in Fig. 2.9 (b), which can capture more photons from multiple facets with a shared optical axis for one point in space. The superposition compound eyes are undoubtedly more found in nocturnal insects. However, some nocturnal bees, ants and wasps have retained the apposition eyes of their day-active ancestors but see remarkably well nonetheless [10]. It implies that the adaptations for nocturnal vision can not only exist in eye structures but also exist in other parts of

visual systems.

2.3.1.2 Dark-adapted Photoreceptors

Photoreceptors respond to photons of light with electrical responses known as bumps [11]. For example, the incoming light is transduced into an electrical signal by photoreceptors, as illustrated in Fig. 2.10. As can be seen, the response amplitude

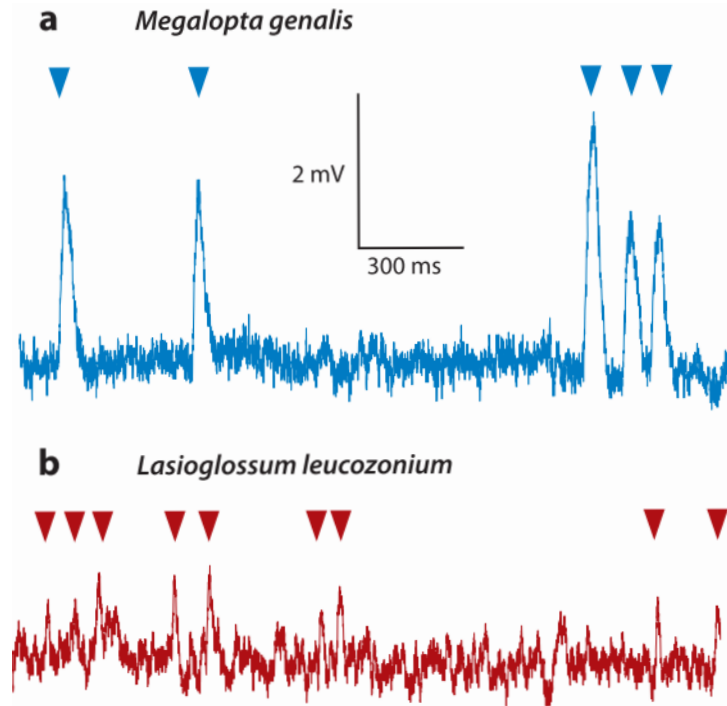


Figure 2.10: Responses to single photons (red and blue arrowheads) recorded from photoreceptors in (a) nocturnal sweat bee (*Megalopta genalis*) and (b) diurnal sweat bee (*Lasioglossum leucozonium*). The figure is adapted from [11].

in Fig. 2.10 varies between the nocturnal bee and the diurnal bee: the bump amplitude is larger, and the bump time course much slower, in *Megalopta* than in *Lasioglossum*. It demonstrates that photoreceptors can increase their response gain when light falls [105]. Moreover, Fig. 2.11 shows locust photoreceptors' voltage response amplitudes relative to the resting potential. It further demonstrates that the voltage response in dark-adapted photoreceptors is different from light-adapted ones.

For photoreceptors modeling, it needs to explain the relationship between photon catch and response amplitude. The 'self-shunting' model [12], [107] can relate the amplitude of potential change at a receptor membrane to the stimulus strength. It used to be widely applied to receptors. The formula of the self-shunting model is given

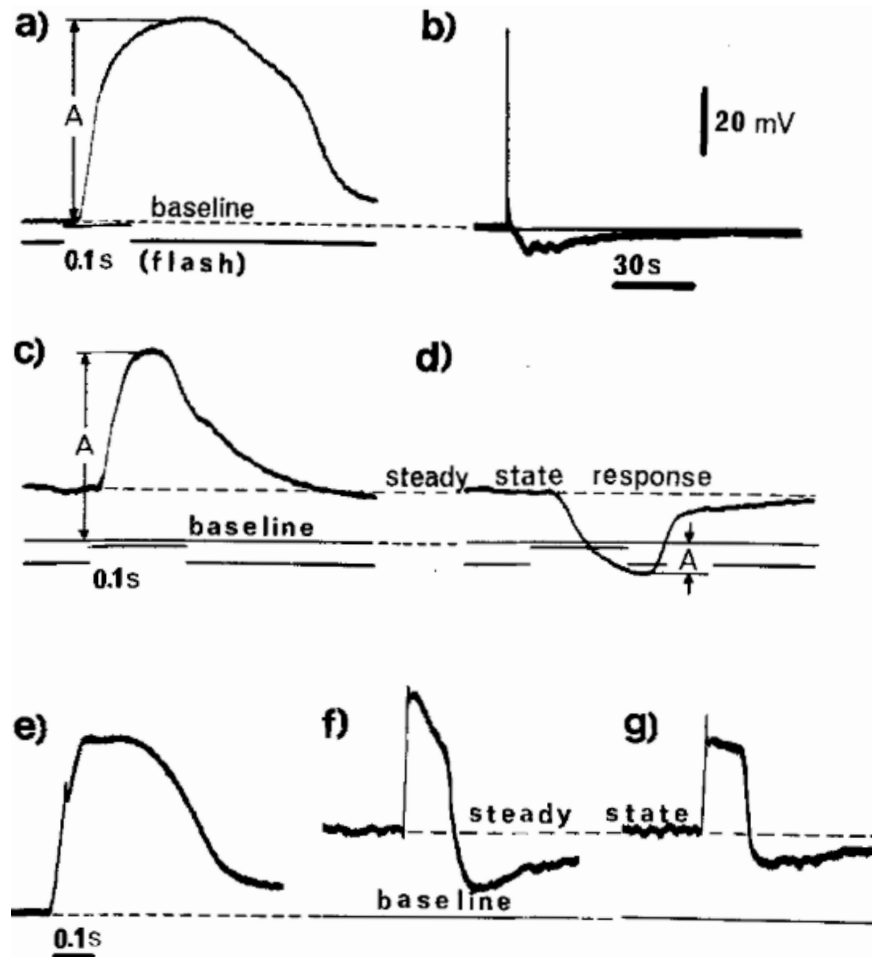


Figure 2.11: The voltage responses of locust photoreceptors. a, b The saturated response of a dark adapted cell recorded on fast and slow time base. c, d Responses of a light adapted cell to an increment above background intensity, and to a brief extinction of the background. e Dark adapted and (f, g) two light adapted responses of different amplitudes. The figure is adapted from [12].

by [12],

$$\frac{V}{V_{max}} = \frac{(RI)^n}{(RI)^n + 1} \quad (2.8)$$

where I is the stimulus intensity, V is the amplitude of the photoresponse, V_{max} is the saturated response amplitude and R is the reciprocal of the intensity yielding a response of 50% V_{max} . The value of power n is related to the receptor organ, which is less than one for many invertebrate photoreceptors. In addition, the power law originated from the famous Stevens' law, which describes expected relationships between perceived intensity and stimulus strength [108]. The general form of the 'power law' can be defined mathematically as follows.

$$\psi = \alpha (I - I_o)^\beta \quad (2.9)$$

where ψ is the perceived intensity, I is the physical intensity, and α , I_o and β are constants.

Shunting networks derive from neurophysiology, and the shunting models of lateral interaction intrinsically are models of adaptation [109]. Together with other means of light adaptation, self-shunting plays an important role in compressing a wide range of light intensities into the narrow functional range of a photoreceptor [110].

2.3.2 Canonical Neural Computations

Physiological and behavioral evidence suggests that the brain relies on a set of canonical neural computations, such as normalization, input filtering, etc. [15], [111]. Normalization as a canonical neural computation exists in a diversity of neural systems in multiple species, from invertebrates to mammals, which can adjust the sensitivity of neurons for performing adaptations [15]. The general normalization model and its applications in the retina are reviewed below.

The general normalization model was defined by a simple equation [15]. That is,

$$R_j = \gamma \frac{D_j^n}{\sigma^n + \sum_k D_k^n} \quad (2.10)$$

where R_j is the normalized response of neuron j , D_j is the driving input (stimulus). The constants, γ , σ and n , are free parameters that fit empirical measurements. In this equation, the normalization factor in the denominator is the sum of a large number of inputs D_k , which is an average pooling operation. In addition, the normalization may involve the nonlinear pooling operation, such as the maximum (MAX) operation [112]. Note that the MAX operation selects the largest responses (winner-take-all), which is able to achieve selectivity and invariance [113].

Fig. 2.12 displays two representative normalization models in the retina and their responses. In Fig. 2.12(a) and (b), the normalization operation of light intensity produces a neural estimate of contrast, and the sensitivity is adjusted by the normalization depending on background light intensity. In Fig. 2.12(c) and (d), the contrast normalization adjusts local contrast, and the effect is probed by gratings that vary in overall contrast and size. The two normalization models in the retina demonstrate that the normalization operation can adjust the gain of neural responses to obtain maximum sensitivity. It is worthy to note that a variety of canonical models have been proposed in sensory systems [114], [115].

2.3.3 Neural Summation

Nocturnal animals are able to see quite well, such as detecting faint movements [116]. Neural summation in the nocturnal animal's visual system extends the limits of vision in dim light [27]. Specifically, summing visual signals in space and time is one strategy to improve visual performance [117]. The anatomical evidence of the neural summation in space is shown in Fig. 2.13. As can be seen, laterally branching visual fibers (L-fibres) are clearly seen in the lamina of the nocturnal bee *Megalopta genalis*, which shows that the spatial summation may take place in the lamina. In addition, nocturnal animals perform slower locomotion at night [116], and this indicates the temporal summation may take place in the visual system.

Warrant [17] proposed a basic summation model (see Fig. 2.14) to explore the effect of neural summation in nocturnal vision. The results revealed that the spatiotemporal summation strategy greatly increased the nocturnal animal's visual reliability.

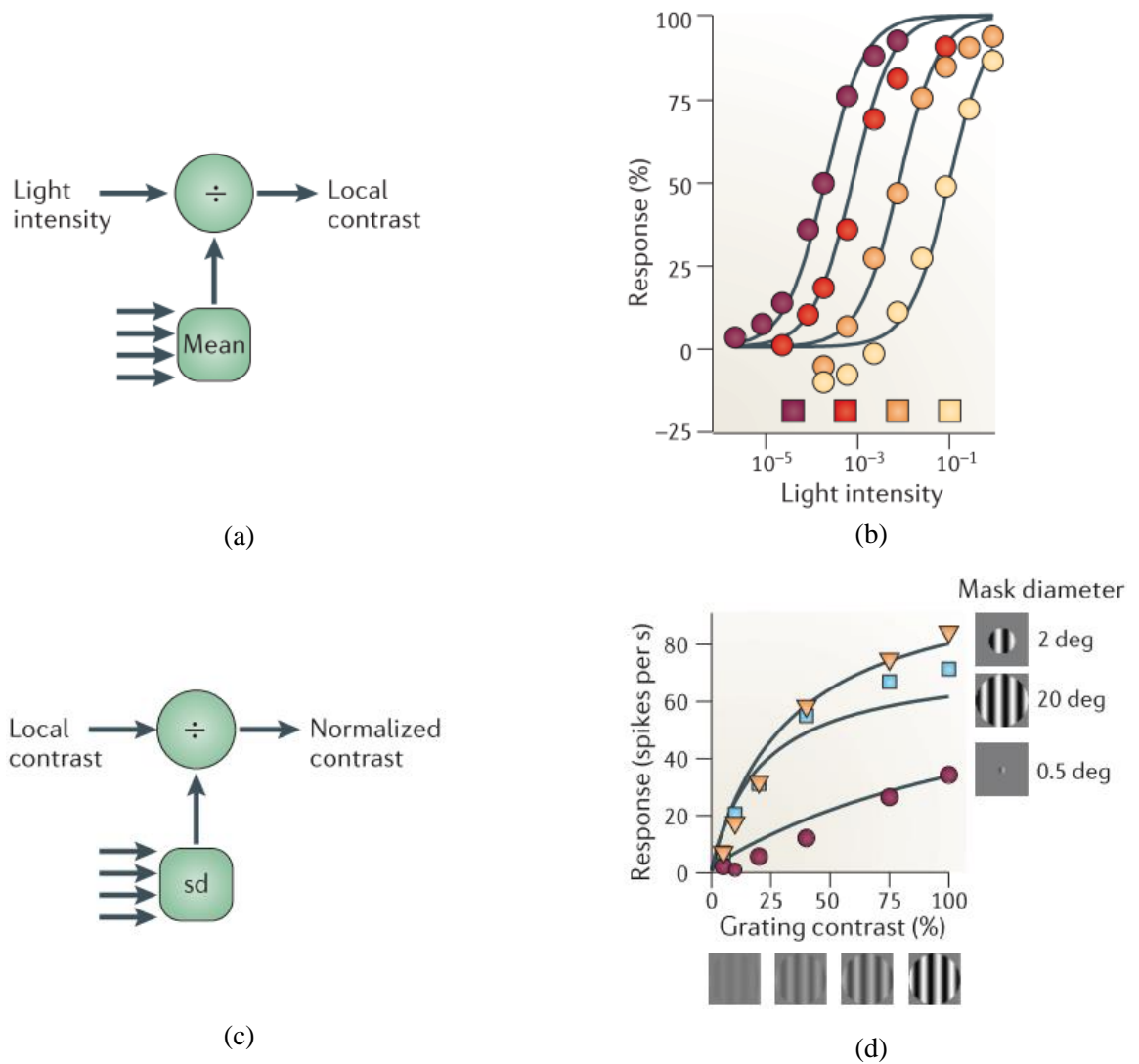


Figure 2.12: Normalization in the retina. (a) Intensity normalization. The multiple arrows indicate light intensities from multiple locations. (b) Responses of intensity normalization, data from [13]. (c) Contrast normalization. (d) Responses of contrast normalization, data from [14]. The figure is adapted from [15].

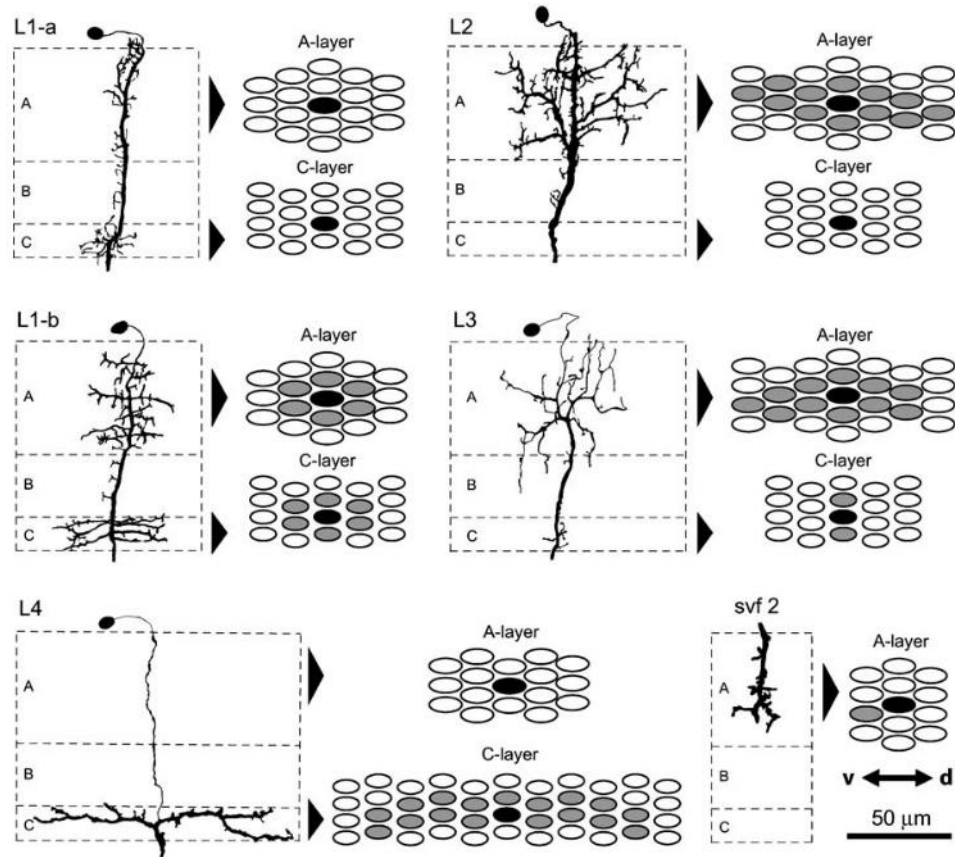


Figure 2.13: Dendritic fields of the short visual fibres (svf 1–3) and first-order interneurons or L-fibres (L1–4) within the first optic ganglion (lamina) of the nocturnal bee *Megalopta genalis*. The black circle indicates the parental cartridge of the fibre, and the neighbouring cartridges reached by its dendrites are shown by grey circles. The lamina has three layers (A, B, and C). Note that v and d indicate ventral and dorsal respectively. The figure is adapted from [16].

Moreover, how spatial and temporal summation achieves the best visual performance was theoretically studied, involving detailed analysis of summation parameters in spatial and temporal domains [117], [118].

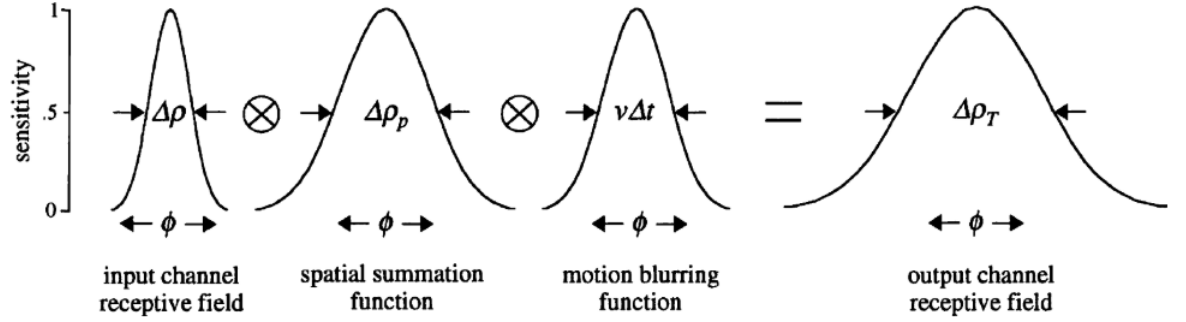


Figure 2.14: Schematic of the basic summation model. The input channel receptive field is modeled as a Gaussian of half-width $\Delta\rho$, and the output channel receptive field is also Gaussian with a half-width $\Delta\rho_T$. In the process of spatiotemporal summation, the spatial summation function has a half-width of $\Delta\rho_p$ degrees and the motion blurring function has a half-width $v\Delta t$ degrees. The function for both the spatial and temporal summation is Gaussian. The figure is adapted from [17].

There were other models proposed to enhance low-light images/videos [11], [42], [116], [119]. These models mainly apply the neural summation strategy for image denoising, as the amplification process of visual signals also amplifies the noise. For example, [11] first adopted the contrast-limited histogram equalization [120] to implement intensity transformation, and then applied neural summation to reduce the amplified noise. However, parameters for summation were determined by smoothing kernels constructed by the structure tensor, which is computationally complex.

The neural summation strategy plays an important role in nocturnal animals' impressive visual abilities. Existing models have verified its effectiveness in enhancing low light images/videos. The detection of faint movements is also the result of the summation strategy [116]. However, how the neural summation function in recognizing a moving object by nocturnal animals remains unclear.

2.4 Low-light Image Enhancement Methods

Over the past several decades, researchers have proposed numerous image/video enhancement methods to improve image quality, including conventional enhancement approaches [19]–[21], [121]–[123] and deep learning-based methods [124]–[127]. Tra-

ditional low-light enhancement methods adopt certain statistical information or assumptions to enhance low-light images. Deep learning-based methods utilize models' powerful learning abilities to achieve low-light image enhancement. In this section, we first review in detail the common histogram-based, Retinex based and gradient-based approaches from the aspect of raising the image intensity, contrast and preserving the structure information. Then, we give a short review of the deep learning-based methods for low-light image enhancement.

2.4.1 Histogram-based Methods

Histogram-based methods [121], [128], [129] adopt the idea of histogram processing to achieve enhancement, which brightens the dark image by adjusting the shape of the histogram. However, histogram processing tends to produce undesirable artifacts since it does not consider the relations between pixels in the image. The details of histogram processing and some histogram-based methods are reported below.

2.4.1.1 Histogram Processing

Histogram processing can enhance low-light images since it provides statistical information on intensities and adjusts the brightness significantly. The computation of histogram of a digital image [130] is given by,

$$h(r_k) = n_k, \quad k = 0, 1, \dots, L - 1 \quad (2.11)$$

$$p(r_k) = \frac{n_k}{M * N} \quad (2.12)$$

where $h(r_k)$ is the histogram of the image, r_k is the k th intensity value, n_k is the number of pixels in the image with intensity r_k , $p(r_k)$ is the normalized histogram. Note that intensity levels of the image histogram is in the range of $[0, L - 1]$ whilst the row and column dimensions of the image are M and N separately.

2.4.1.2 Histogram Equalization Method

Histogram equalization (HE) [130] is a popular image enhancement method due to its simplicity and efficiency. The basic idea of HE is to modify the intensity distribution to a uniform histogram. Intensity mappings of HE can be defined by

$$s = T(r) = (L - 1) \sum_{j=0}^r p_r(j), \quad 0 \leq r \leq L - 1 \quad (2.13)$$

where j is a dummy variable, $p_r(j)$ denotes the probability density function (PDF) of j . Note that the value of $p_r(j)$ is calculated by Eq. 2.12. The HE method relies on the probability density function (PDF) of pixels of the input image and treats pixels in the image individually. Thus, the HE often produces saturated outputs like very bright or very dark intensities.

2.4.1.3 Adaptive Histogram Equalization Method

The adaptive histogram equalization (AHE) and its variants [120] were designed to overcome some concerns. For example, the contrast limited adaptive histogram equalization (CLAHE) method can effectively enhance image contrast by limiting the extent of contrast enhancement [11]. However, the output image usually contains many noises when the input has mixed bright and dark regions, as the clipping histogram process in CLAHE only relies on the limitation parameter.

2.4.2 Retinex-based Methods

Retinex, coined by Land [131], is a valid theory aiming at simulating the human visual system (HVS) to achieve color constancy and dynamic range compression [122]. The Retinex theory based on the physical imaging model can be mathematically described by

$$I(x, y) = R(x, y) L(x, y) \quad (2.14)$$

where $R(x, y)$ is the reflectance of the imaged object and $L(x, y)$ is determined by the illumination source. Retinex-based methods enhance the image by calculating the illumination of the image and removing them.

Single-scale Retinex (SSR) [132] and multi-scale Retinex (MSR) [133] are two representative methods that apply the Retinex theory to enhance images. The SSR and the MSR take the reflectance as the final enhanced result by applying local Gaussian filters to remove illumination [134]. However, their results often look unnatural and frequently appear to be over-enhanced [20]. In recent years, some new Retinex-based methods [20], [135], [136] were proposed to achieve better enhancement results by more accurate estimation of illumination map and reflectance. Although these new methods are efficient, they heavily depend on certain assumptions.

2.4.3 Gradient-based Methods

Since the assumption [137] that the human visual system (HVS) is more sensitive to the gradient than to the absolute luminance reaching the retina, some gradient-based methods [21], [123], [138] were proposed to enhance images. These methods apply gradient information to improve the visual performance of dark regions. For example, [138] brings out detail in dark regions by attenuating the magnitudes of large gradients, and [21] integrates the enhanced gradients of the dark with the intensity-range constraint to highlight details in low-light regions.

Gradient-based methods can effectively preserve the edge information while enhancing low-light images. However, enhancement performance relies on an appropriate manipulation or integration of the gradient and the intensity. In complex dim-light scenes, these gradient-based methods easily produce undesirable results.

2.4.4 Deep Learning-based Methods

Recently, some methods were proposed to enhance low-light images by taking advantage of the latest deep learning technology [125]–[127], [139]. The key to the deep learning-based methods is to learn a generic enhancement mapping through the training, which requires the dataset of low-light and well-lighted image/video pairs. However, capturing the pairs of real-world low-light images and bright ones (ground truth) is difficult. Although the synthesis strategies in [126], [139] and designs for the imaging system in [125], [127] can help to produce training and ground truth pairs,

there are still limitations for dataset capacity to reflect highly diverse scenes and objects. Moreover, the data-driven low light image enhancement approaches usually have high complexity networks and thus are computationally complex, which may not be suitable for real-time applications like motion detection.

2.5 Motion Cues Detection Methods

Motion cues by luminance-dependant visual processing play an important role in guidance and navigation in the insect visuomotor system [140], [141]. The algorithms and neural circuits for the extraction of visual motion cues have been the focus of intense research [142]. In this thesis, we regard the edges of moving objects as motion cues. Motion cues detection methods are generally categorized into temporal differencing, optical flow and motion detectors.

2.5.1 Temporal Differencing

Motion cues can be extracted by detecting regions of change in multiple images of the same scene taken at different times [143]. Temporal differencing, also named frame differencing, is the simplest method for motion cues detection, which calculates the pixel-wise differences between two consecutive frames [18], [144], [145]. It is mathematically defined by,

$$I_D(x, y, t) = |I(x, y, t) - I(x, y, t - \Delta t)| \quad (2.15)$$

where $I(x, y, t)$ and $I(x, y, t - \Delta t)$ represent images captured at the current frame and the previous frame separately, $I_D(x, y, t)$ denotes the absolute difference between $I(x, y, t)$ and $I(x, y, t - \Delta t)$. Note that Δt indicates the frame interval as continuous time is discretized by frames. When the difference image $I_D(x, y, t)$ is compared with a given threshold, a roughly moving area containing motion cues is obtained. Fig. 2.15 displays the motion cues on the result of frame differencing.

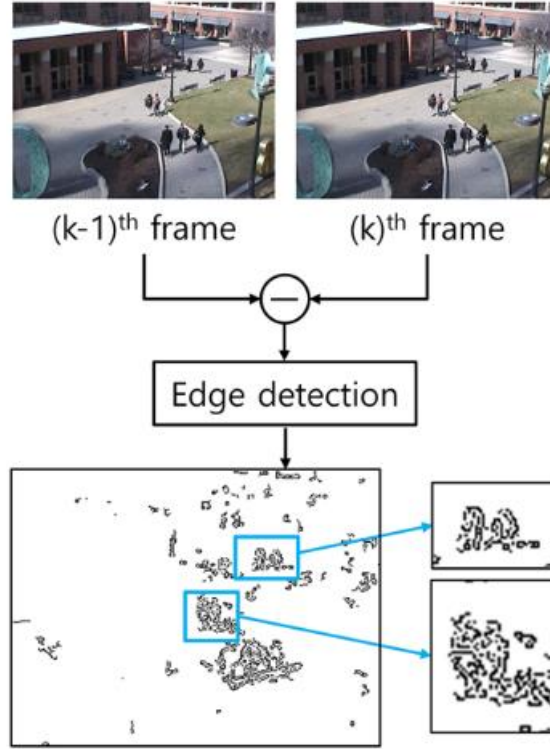


Figure 2.15: Motion cues on the result of frame differencing. The figure is adapted from [18].

2.5.2 Optical Flow

Optical flow methods can also be used to extract motion boundaries (edges) for tracking moving objects, e.g. [146]–[148]. The motion cues are obtained based on the computation of optical flow between pairs of subsequent frames. Several methods for the computation of the optical flow have been introduced in subsection 2.2.1.1. According to [147], optical flow algorithms that can support large displacements between frames are applicable, such as the algorithm [149]–[151]. The detailed process of estimating motion boundaries in [147] is described in the following. First, the optical flow between two consecutive frames with respect to pixel p is denoted by \vec{f}_p . Then, three estimators are used to compute motion boundaries. The first estimator is based on the magnitude of the gradient of the optical flow. That is,

$$b_p^m = 1 - \exp\left(-\lambda^m \left\| \Delta \vec{f}_p \right\| \right) \quad (2.16)$$

where b_p^m denotes the strength of motion boundaries at pixel p , which is in the range of $[0, 1]$. And, the λ^m is a controlling parameter that controls the steepness of the

function. The second estimator is based on the direction between the motion of pixel p and its neighbors \mathcal{N} . That is,

$$b_p^\theta = 1 - \exp\left(-\lambda^\theta \max_{q \in \mathcal{N}} (\delta\theta_{p,q}^2)\right) \quad (2.17)$$

where $\delta\theta_{p,q}$ denotes the angle between \vec{f}_p and \vec{f}_q , which is in the range of $[0, 1]$. The third estimator is based on the combination between the first estimator b_p^m and the second estimator b_p^θ . That is,

$$b_p = \begin{cases} b_p^m, & \text{if } b_p^m > T \\ b_p^m \cdot b_p^\theta, & \text{if } b_p^m \leq T \end{cases} \quad (2.18)$$

where T is a high threshold, and a binary motion boundary can be determined and labeled. Fig. 2.16 exhibits the results of motion boundaries from the aforementioned three estimators.

2.5.3 Motion Detectors

In insects, there are various motion-sensitive neurons that can respond vigorously to visual stimuli with specific directions, e.g. direction-selective neurons (DSNs) [152] and lobula giant movement detectors (LGMDs) [24]. Specifically, these motion-sensitive neurons are excited by preferred direction stimuli which are luminance changes in spatio-temporal domain caused by moving objects. Also, the moving edges can be regarded as visual motion stimuli.

The motion detector model the motion sensitive neuron in biological visual systems, which respond to moving edges for perceiving visual motion via excitatory and inhibitory signals processing in specific ways [153]. For example, a lot of bioplausible neural network models were proposed to mimic response properties of the motion-sensitive neurons, such as the four whole-field direction-selective neural networks (DSNNs) [6], [51], [153], [154] and the LGMD1 neural network [7], [41], [155], [156]. These motion detectors can also be integrated, as illustrated in Fig. 2.17 (a) and (b), for specific applications. It is worthy to note that the accurate extraction of motion

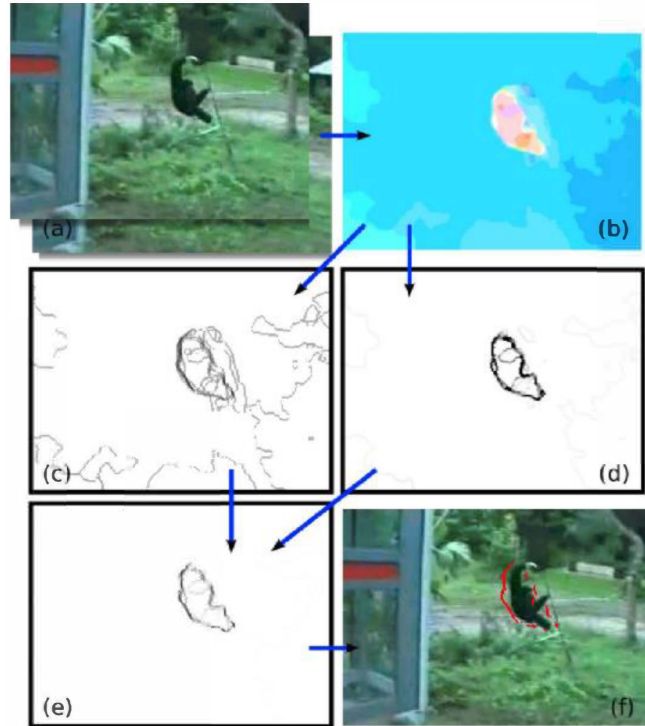


Figure 2.16: Motion boundaries of moving objects by the computation of optical flow. (a) Two input frames. (b) Optical flow \vec{f}_p . (c) Motion boundaries, based on the magnitude of the gradient of the optical flow b_p^m . (d) Motion boundaries, based on difference in direction between a pixel and its neighbors b_p^θ . (e) Combined motion boundaries b_p . (f) Final, binary motion boundaries after thresholding, overlaid on the first frame. The figure is adapted from [18].

cues is critical for these neural network models to work robustly. Fig. 2.17 displays motion cues resulting from specific layers in the motion detectors.

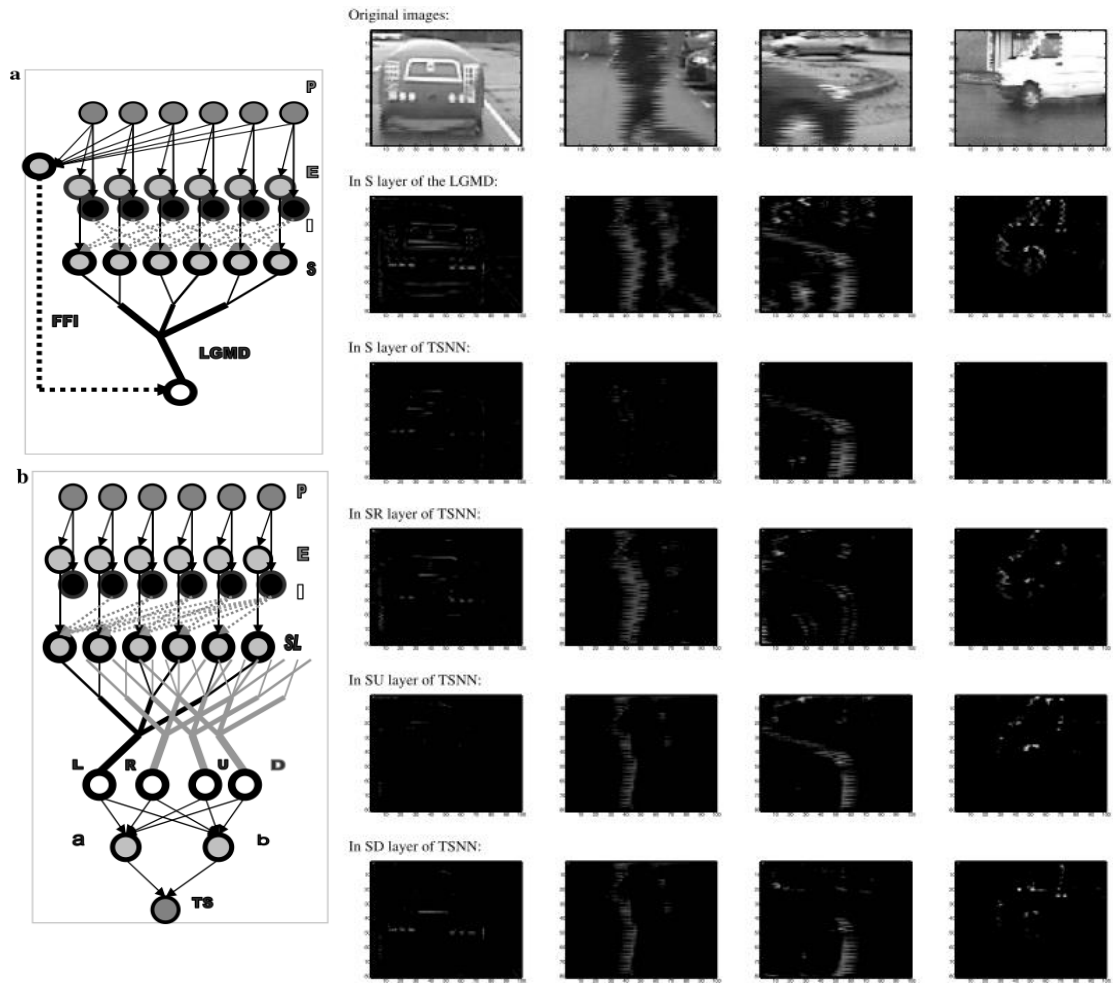


Figure 2.17: Motion cues resulting from specific layers of LGMD1 and TSNNs. a. LGMD1 neural network model. b. TSNN neural network model, which includes four DSNNs (L-Left, R-Right, U-Up, D-Down). The figure is adapted from [6].

Chapter 3

LGMD1- Looming Cue Detection Against Translating Motion

Insects have evolved exquisite visual systems to avoid obstacles or detect approaching predators in complex visual scenes for better survival [23]–[25]. Their compacted brains and specialized neural mechanisms become a source of inspiration for autonomous robots and vehicles in specific functionalities such as navigation, tracking, collision detection, etc [26], [39]–[41], [157]. LGMD1 neuron is one of the lobula giant movement detectors [29], [158] in the locusts’ brain, which plays an important role in protecting the locust from a collision. It responds selectively to looming objects when compared with non-looming stimuli like translation. However, the ability to robustly inhibit translation motion but only responds strongly to looming objects is still a challenge for the existing LGMD1-based artificial visual systems [7], [8], [30], [31]. Hence, how neural strategies shape the LGMD1’s selectivity effectively for looming forms the basis of the present research.

The selective response property requires the system only produces large output in the preferred direction [154], [159]. It is well-known that the looming motion is quite different from translation. The former is characterized by an expanding size of motion image conveying information about movement in depth [160]. The latter exhibits the change of position of moving images in the field of view without the change of size. Although existing LGMD1 models respond strongly to looming objects, the fast, large translating object or the object at changing speed may also elicit similar responses. It

becomes difficult to distinguish looming and translating motion under such circumstances. How neural processing can inhibit the translatory stimuli effectively is essential in constructing a new LGMD1-based visual neural network that only responds to looming cues.

Neural information processing, like lateral inhibition [5], spiking frequency adaptation (SFA) [33], separated processing of ON/OFF signals [8], and synaptic plasticity mechanisms [26] are applied in previous models. But they are not ideal methods when dealing with challenging translating stimuli, for example, translatory objects at loom-like speed or moving in a divergent mode. By analyzing the response properties of two kinds of motion patterns in spatial and temporal domains, we found the translating object usually triggers paired ON-OFF responses while looming only produces monotonous responses either ON or OFF. It implies a neural competition may occur in the LGMD1 neuron.

The neural competition can help select some certain response feature or pattern [34], [35]. Moreover, the phenomenon of 'canceling' each other between ON and OFF edges is also observed in the mouse retina [61] and similar models are proposed for detecting approaching objects [64], [65]. In the postsynaptic partner of LGMD, the well-known DCMD neuron, a clear reciprocal relationship between ON and OFF pathways can be found [161]. Therefore, it is possible to propose a new bio-plausible LGMD1 model by comparing ON and OFF responses. The new computational structure of LGMD1 is based on the framework of separated ON/OFF pathways, incorporating neural information processing mechanisms. The key functionality of neural competition is to filter out translating stimuli. Additionally, we add a new denoising process in the model inspired by the role of lateral excitation [67] and the group-decay processing mechanism [7]. The process is responsible for extracting real collision information from backgrounds with noise, which makes the looming detection much more reliable.

In this chapter, we propose an LGMD1 neural network model that shows a strong direction preference between looming and translation motion. The proposed model is sensitive to looming objects but robust to inhibiting translatory objects. The original contributions of this work are: 1) a new computational architecture is proposed

based on the neural competition between ON and OFF pathways; 2) the denoising mechanism is implemented for extracting reliable motion cues; 3) more comprehensive comparative experiments for looming and translating events are conducted and the improved results report that the proposed model shows robustness in detecting looming object against translation. The rest of this chapter is organized as follows. In Section 3.1, we introduce the network architecture of the proposed LGMD1 model. Section 3.2 provides comprehensive experimental results, including performance evaluation and comparisons against baseline models. Section 3.3 finally presents the conclusion and future work.

3.1 Formulation of the LGMD1 Model

The proposed LGMD1 model is shown in Fig. 3.1. The model separates the ON and OFF channels for processing visual signals. In an advance on the former LGMD1 model [9], it incorporates new D_E and D_I layers before the S layer and removes the previous G layer. In the following subsections, we have schematically illustrated the motion signal processing architecture of the proposed model and elaborate on its components.

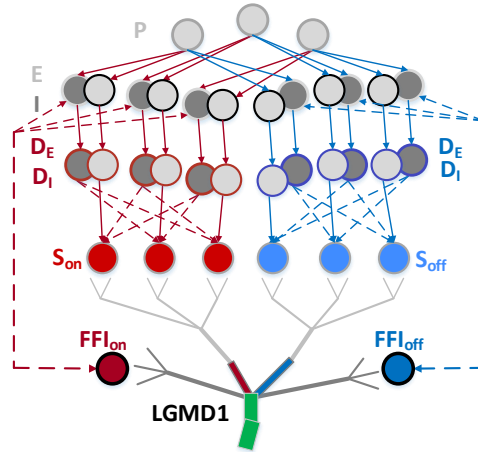


Figure 3.1: The proposed LGMD1 model is composed of six layers of cells (P, E, I, D_E , D_I , S) and three single cells (FFI_{on}, FFI_{off}, LGMD1). Signals are split into ON (red-arrows) and OFF (blue-arrows) channels each with five layers (E, I, D_E , D_I and S); the D_E and D_I layers are responsible for filtering out isolated background excitation. The dashed lines indicate transmission of delayed neural signals. Excitatory signals from the two channels create neural competition in the LGMD1 neuron.

Fig.3.2 shows the architecture of the network, including the retina, lamina, medulla,

and lobula layers. ON or OFF type stimuli represent responses of positive or negative contrast polarity elicited by moving edges. The two different types of stimuli are then further processed in separate ON and OFF channels. More details are described as follows.

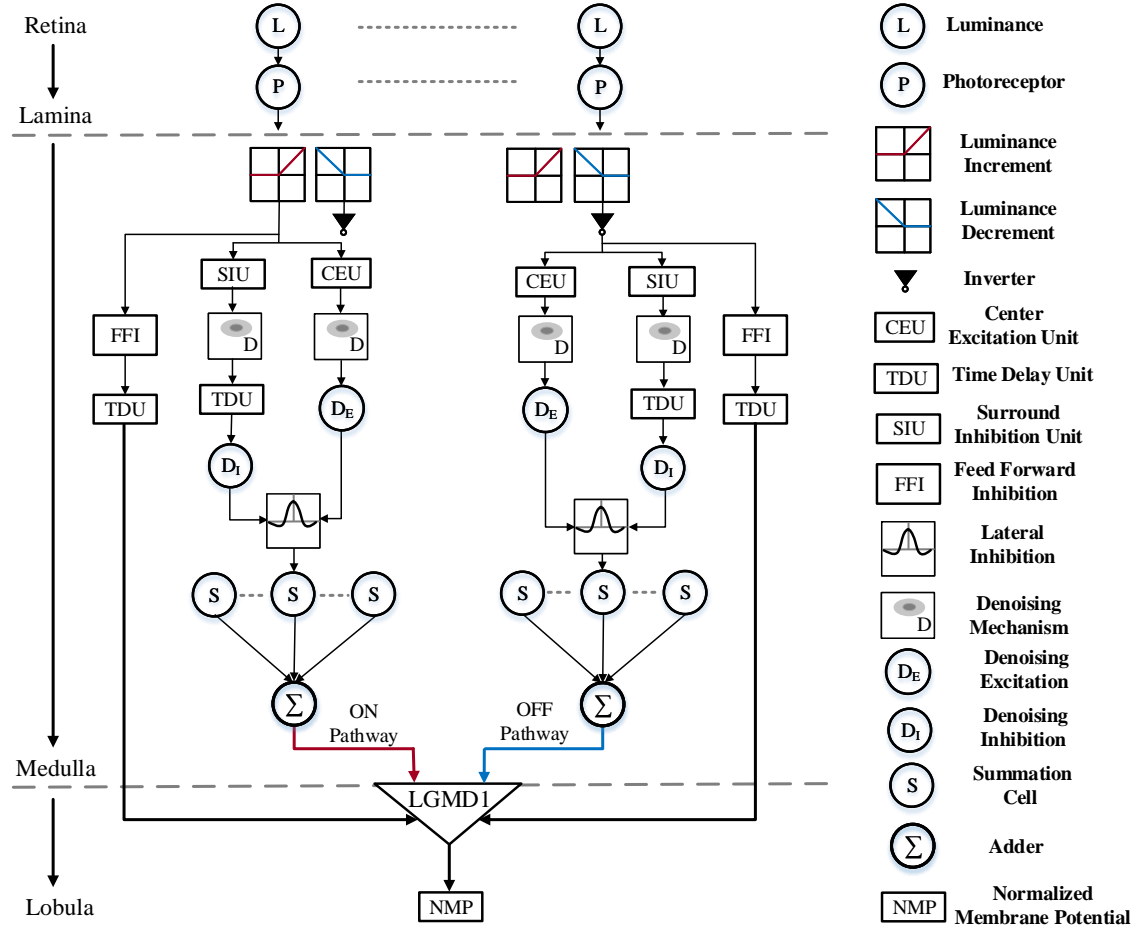


Figure 3.2: Schematic illustration of the three stages of signal processing in the LGMD1 model. Retina to lamina layer: the photoreceptor (P) cell captures changes in pixel-wise luminance (L); each P-unit feeds both an ON and an OFF channel. Lamina to medulla layer: two types of visual stimuli are processed separately in ON and OFF channels. This includes an asymmetric mechanism (rectification), a denoising mechanism (see Fig. 3.3) and a lateral inhibition mechanism (see Fig. 3.4). Medulla to lobula layer: the LGMD1 neuron collects the winner excitatory signals arising from the neural competition between ON and OFF pathways; the delayed feed forward inhibition (FFI) signals from the ON and OFF responses inhibit the initial response of the model to movement or sudden great changes in luminance in the visual field.

3.1.1 Retina to Lamina layer

The photoreceptor (P) cells detect changes in luminance (L) and produce corresponding electrical signals. The output of the P cell is defined by the following equa-

tions:

$$P(x, y, t) = \frac{1}{\int_{t-T}^t p_i di} \cdot \int_{t-T}^t p_i \cdot L'(x, y, i) di \quad (3.1)$$

$$p_i = e^{-\alpha \cdot (t-i)}, p_i \in (0, 1) \quad (3.2)$$

$$L'(x, y, i) = L(x, y, i) - L(x, y, i - \Delta i) \quad (3.3)$$

where $P(x, y, t)$ reflects the change in luminance detected by pixel (x, y) over time T milliseconds. This indicates the persistence time that the change in luminance can last. p_i represents the attenuation coefficient of L' and the parameter α indicates its attenuation rate. The luminance change $L'(x, y, i)$ is computed by Equ. (3.3), where $L(x, y, i)$ and $L(x, y, i - \Delta i)$ indicate the gray values of pixel (x, y) at time i and $i - \Delta i$.

The P cells are arranged in a matrix mapping pixel positions in video images. Moreover, continuous-time is discretized by a sequence of image processing frames. Time i and $i - \Delta i$ indicate two consecutive frames with a frame interval Δi . Note that the value of T is set empirically by three frame intervals, and its value varies when the frame rate changes.

3.1.2 Lamina to Medulla layer

The changes in luminance (increments or decrements) are first rectified by transient cells in the medulla [162], as shown in Fig. 3.2. Then, signals with opposite polarity go through three processing stages in separate ON and OFF pathways. Each channel exhibits a similar computational process, which is elaborated as follows.

I) Asymmetric Mechanism: A pairwise excitatory (E) and inhibitory (I) cells receive the rectified outputs from the corresponding P cell. The ‘half-wave rectification’ [142] processing converts the polarity of negative (OFF) into positive, but the positive (ON) value remains unchanged. The rectification process for the ON and OFF channels is given by,

$$E_{\text{ON}}(x, y, t) = I_{\text{ON}}(x, y, t) = [P(x, y, t)]^+ \quad (3.4)$$

$$E_{\text{OFF}}(x, y, t) = I_{\text{OFF}}(x, y, t) = -[P(x, y, t)]^- \quad (3.5)$$

where $[a]^+ = \max(0, a)$, $[a]^- = \min(0, a)$. The output of the E/I cell is related with image contrast, which have the same value.

2) Denoising Mechanism: The denoising mechanism aims to extract the real stimuli and eliminate small and isolated background excitations. The process in the E/I layer can be interpreted by Fig. 3.3 and consists of two stages: first, computing the value of the passing coefficient of each cell; then transmitting the value of the signal from each cell multiplied by corresponding passing coefficient to the D_E/D_I layer.

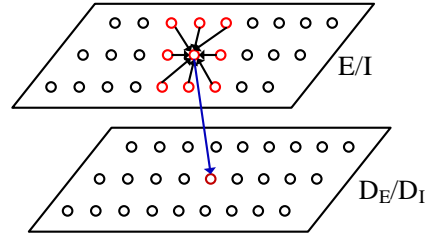


Figure 3.3: Schematic illustration of the denoising mechanism. The red cells in E/I layer represent the receptive field of the cell (red) in the D_E/D_I layer. Two processing stages are implemented between the two layers, including the computation of the passing coefficient (the black arrows) and the transfer of the expected excitation or inhibition signal (the blue arrow).

In the first stage, the passing coefficient is related to the average change in luminance of the receptive field and it is normalized by dividing by the maximum average luminance change in the whole layer. According to [7], the receptive field is the cell's surrounding neighbors and the neighbors' outputs are transmitted by a convolution process using a connection weighting matrix. Note that the connection weighting matrix of the receptive field in the E/I layer can be represented by a 3×3 matrix,

$$W_E = W_I = \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix} \quad (3.6)$$

where W_E and W_I are convolved with the excitation and inhibition matrix in ON and OFF pathways. The average luminance change (excitation or inhibition) can be math-

ematically defined as:

$$A_E(x, y, t) = \sum_{i=-1}^1 \sum_{j=-1}^1 E(x+i, y+j, t) W_E(i, j) \quad (3.7)$$

$$A_I(x, y, t) = \sum_{i=-1}^1 \sum_{j=-1}^1 I(x+i, y+j, t) W_I(i, j) \quad (3.8)$$

where $E(x, y, t)$ and $I(x, y, t)$ represent the excitation and inhibition in the receptive field of each cell. The normalized value of E/I cell can also be regarded as the probability of occurrence of luminance changes,

$$P_E(x, y, t) = A_E(x, y, t) / (\Delta c + \max(A_E)) \quad (3.9)$$

$$P_I(x, y, t) = A_I(x, y, t) / (\Delta c + \max(A_I)) \quad (3.10)$$

where $P_E(x, y, t)$ and $P_I(x, y, t)$ are normalized passing coefficients, $\max(A_E)$ and $\max(A_I)$ denote the range of luminance changes. Δc is a small real number to prevent the denominator tending toward zero.

In the second stage, we pass the expected excitation and inhibition signals through two separate channels by multiplying the excitation value by the normalized passing coefficient,

$$D_E(x, y, t) = P_E(x, y, t) \cdot E(x, y, t) \quad (3.11)$$

$$D_I(x, y, t) = P_I(x, y, t) \cdot I(x, y, t) \quad (3.12)$$

where $D_E(x, y, t)$ or $D_I(x, y, t)$ denote the denoising cells. They become excited only when their value exceeds the threshold T_c . This can be mathematically described as follows.

$$\tilde{D}_E(x, y, t) = \begin{cases} D_E(x, y, t) & \text{if } D_E(x, y, t) \geq T_c \\ 0 & \text{otherwise} \end{cases} \quad (3.13)$$

$$\tilde{D}_I(x, y, t) = \begin{cases} D_I(x, y, t) & \text{if } D_I(x, y, t) \geq T_c \\ 0 & \text{otherwise} \end{cases} \quad (3.14)$$

where T_c represents the threshold of the change in luminance. This process is capable of retaining the strong signals whilst eliminating small or isolated signals in the background. Since the excitation and inhibition have the same value (see Equ. (3.4) and (3.5)), the threshold for D_E and D_I are identical.

3) Lateral Inhibition Mechanism: In the early stages of visual processing, lateral inhibition is a very common feature in many organisms [163]. The detail of the lateral inhibition mechanism is illustrated in Fig. 3.4. They are mathematically defined as follows.

$$L_{I_{ON}}(x, y, t) = \sum_{i=-1}^1 \sum_{j=-1}^1 \tilde{D}_{I,ON}(x+i, y+j, t-\tau_s) W_i(i, j) \quad (3.15)$$

$$S_{ON}(x, y, t) = [\tilde{D}_{E,ON}(x, y, t) - L_{I_{ON}}(x, y, t)]^+ \quad (3.16)$$

$$\tilde{S}_{ON}(x, y, t) = \lambda \cdot \log_{10}(S_{ON}) \quad (3.17)$$

where $L_{I_{D,ON}}(x, y, t)$ represents the summation of lateral inhibitions from the D layer corresponding to a spatial position (x, y) at time t in the ON pathway. $\tilde{D}_{I,ON}$ and $\tilde{D}_{E,ON}$ represent denoising inhibitions and excitations in the ON channel, which is summated by the S cell. τ_s represents the time delay constant and is set to one frame delay in this model. W_i indicates the connection weight matrix of lateral inhibition (see Fig. 3.4(b)). $S_{ON}(x, y, t)$ denotes the output of S cell. \tilde{S}_{ON} indicates the normalization value of S_{ON} that can deal with low-contrast situations. λ controls the expected output value of the S cell via the spiking threshold of the LGMD1 neuron which is determined empirically. The OFF pathway implements a similar processing procedure to the ON pathway.

3.1.3 Medulla to Lobula layer

In the lobula layer, numerous trans-medullary-afferents (TmAs) connect the eye with the LGMD1 neuron [24]. The excitatory signals from the S cells are passed by TmAs to the LGMD1 neuron. A neural competition between summed ON and OFF type responses is schematically illustrated in Fig. 3.5. If the membrane potential

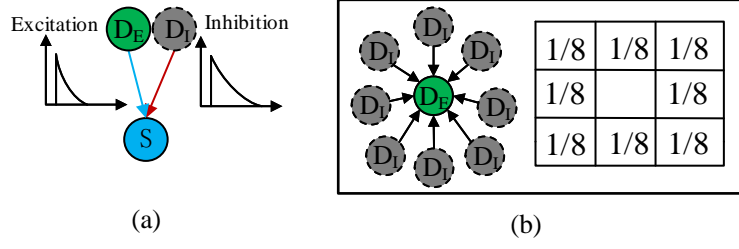


Figure 3.4: Schematic illustration of the lateral inhibition mechanism. (a) The excitation (D_E) unit and the neighboring inhibition (D_I) unit from the D layer are summed to the summation (S) unit in the same retinotopic position; Excitation and inhibition decay exponentially with different coefficients, and the decay processes are described in [5]; Delays at excitatory connections are set to 0 milliseconds, whereas delays at inhibitory connections are set to vary from several to tens of milliseconds accordingly (one frame delay). (b) Connection weight matrix of lateral inhibition.

evoked by winner excitations exceeds a fixed threshold, it will elicit a spike. The computational process of the neural competition is depicted as follows.

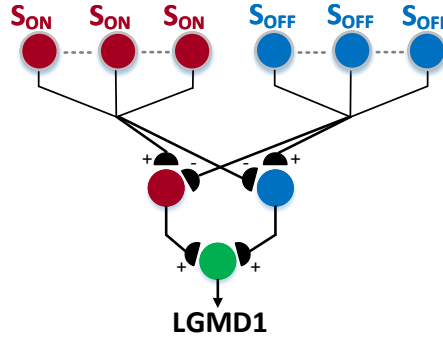


Figure 3.5: Schematic illustration of the neural competition between ON and OFF channels. S_{ON} and S_{OFF} are effective outputs (nonzero value) from the S layer participating in competition. The winner excitations are passed to the LGMD1 neuron.

Neural Competition Mechanism: The summed outputs of S cells in separate ON and OFF channels make a response comparison. That is,

$$S_{-}\tilde{S}_{ON}(t) = \sum_x \sum_y \tilde{S}_{ON}(x, y, t) \quad (3.18)$$

$$S_{-}\tilde{S}_{OFF}(t) = \sum_x \sum_y \tilde{S}_{OFF}(x, y, t) \quad (3.19)$$

where $S_{-}\tilde{S}_{ON}(t)$ and $S_{-}\tilde{S}_{OFF}(t)$ indicate the summation of nonzero excitations in ON and OFF pathways respectively. The winner excitations from the comparison between two opponent pathways represent the practical expanding stimuli. Correspondingly,

the value of membrane potential $MP(t)$ is defined as:

$$C_{max}(t) = \max(S_{-}\tilde{S}_{ON}(t), S_{-}\tilde{S}_{OFF}(t)) \quad (3.20)$$

$$C_{min}(t) = \min(S_{-}\tilde{S}_{ON}(t), S_{-}\tilde{S}_{OFF}(t)) \quad (3.21)$$

$$MP(t) = \begin{cases} C_{max}(t) - C_{min}(t) & \text{if } C_{min}(t) \leq 1 \\ C_{max}(t) / C_{min}(t) - 1 & \text{if } C_{min}(t) > 1 \end{cases} \quad (3.22)$$

where the output of $MP(t)$ is scaled when the value of C_{min} is larger than 1. The membrane potential is then normalized by an activation function. That is,

$$NMP(t) = 1 - (1/\exp(\beta^{-1} \cdot MP(t) \cdot n_{cell}^{-1})) \quad (3.23)$$

where β is the parameter related to the angular size (θ_{thres}) on the locust's retina, and n_{cell} denotes the total number of cells (numbers of pixels) in the G layer, $NMP(t) \in (0 \sim 1)$. It is worthy to note that the peak firing rate always occurs after the object has reached a threshold angle θ_{thres} of $24^\circ (\pm 1.5^\circ)$ on the locust's retina [62]. Hence, β is set to 0.25 when the field of view (FOV) of the camera eye is nearly 100° .

3.1.4 Spiking Mechanism

If the value of $NMP(t)$ exceeds the threshold T_s , a spike is produced,

$$Spike(t) = \begin{cases} 1 & \text{if } NMP(t) \geq T_s \\ 0 & \text{otherwise} \end{cases} \quad (3.24)$$

where 1 represents a spike, 0 indicates no spike. The alarm time (A_T) represents the occurrence of successive n_{sp} spikes in a specific time period t_n . That is,

$$A_T = \begin{cases} t & \text{if } \int_{t-t_n}^t Spike(i) \geq n_{sp} \\ 0 & \text{otherwise} \end{cases} \quad (3.25)$$

where the value of A_T is greater than ZERO indicates a collision. The alarm frame (A_F) is the discretized format of A_T . Note that the value of T_s and n_{sp} are set in accordance with [7].

3.1.5 The Feed Forward Inhibition (FFI)

The FFI is said to have the ability to suppress large changes in the image caused by ego-motion [30] or the model's initial response to movement [163]. However, the FFI is transiently activated in response to large and rapid changes in luminance and is therefore unlikely to contribute to the sustained suppression of the responses to translating objects [78]. In this chapter, the response to translational motion caused by ego-motion can be effectively inhibited by the ON/OFF neural competition mechanism (see Section 3.2.4). Hence, FFI is simply used to shut down the spiking of LGMD1 when substantial changes in luminance occur in the visual scene caused by changes in ambient lighting conditions. The FFI signal is gathered from P cells with a time delay, which is mathematically described by,

$$FFI_{ON}(t) = \sum_x \sum_y P_{ON}(x, y, t - \tau) \quad (3.26)$$

$$FFI_{OFF}(t) = \sum_x \sum_y P_{OFF}(x, y, t - \tau) \quad (3.27)$$

$$FFI(t) = FFI_{ON}(t) + FFI_{OFF}(t) \quad (3.28)$$

where $FFI(t)$ represents the changes in luminance received from the photoreceptors with a delay of τ milliseconds from both ON and OFF channels. The value of τ is set to one frame interval. If the value of FFI exceeds a threshold, the spikes in the LGMD are inhibited immediately. It should be noted that the FFI has not been activated in the experiments that have been carried out to validate the inhibitory effect of neural competition to translational motion caused by ego-motion.

3.1.6 Parameters of the System

Parameters of the proposed LGMD1 model are listed in Table 3.1, based on current trials. Certain parameters (Δc , β and n_{cell}) have been determined by the analyses in Sections 3.1.1 ~ 3.1.5. Other parameters have been chosen empirically. Note that the value of parameters T , τ_s and τ are adjusted by the frame rates of the sequences tested (30 ~ 120 fps). For example, T (25 ms), τ (8.3 ms) and τ_s (8.3 ms) when the frame rate equals 120 fps. In the following experiments, they are kept unchanged unless stated.

The proposed LGMD1 model is written in Matlab (The MathWorks, Inc., Natick, MA). The computer used in experiments is a standard laptop with a 1.80GHz Intel Core i5 CPU and 16.00GB RAM memory.

Table 3.1: Parameters of the proposed LGMD1 model

Name	Value	Name	Value	Name	Value
T	25 ~ 100 ms	α	0.04	Δc	0.01
T_c	10	λ	50	β	0.25
T_s	0.7	n_{sp}	4	n_{cell}	12288
τ_s	8.3 ~ 33.3 ms	τ	8.3 ~ 33.3 ms		

3.2 Experimental Results

3.2.1 Experimental Setup

1) Data Sets: The proposed model was tested on both synthetic and real datasets. The synthetic image sequences were produced by Vision Egg software [164], including traditional sinusoidal gratings, the movement of simple shapes on clean or real complex backgrounds (see Section 3.2.2, Section 3.2.3, Section 3.2.5) as well as the shifting of a panoramic natural scene (see Section 3.2.4). The real image sequences were recorded with a GoPro camera which was fixed on a tripod in front of simple and complex environments, as well as fixed on a mobile Yahboom 4WD smart robot in a complex environment (see Section 3.2.5). The GoPro camera was set with a wide field of view (FOV) mode such that the vertical and horizontal FOV were 94.4° and 122.6° respectively. The resolution of the synthesized image was 96×128 pixels and

the frame rate ranged from $30 \sim 120$ frames per second (fps). The captured real image had a resolution of 960×1280 pixels at 120 fps, but the image that was fed as an input to the neural network models was resized to 96×128 pixels using the ‘imresize’ function in the Matlab image processing toolbox. All visual stimuli in the data sets can be divided into looming and translating events.

2) Evaluation Criteria: We employ the normalized membrane potential (NMP) curve to show the output of LGMD1 models. If the result of an LGMD1 responding to visual stimuli is correct, it indicates a successful detection. The result is related to the angular threshold (θ_{thres}), time to collision (T_C) or alarm frame (A_F). The value of θ_{thres} is independent of the size or velocity of the approaching object. According to previous biological research [62], the peak LGMD1 activity occurs after the approaching object has reached a specific angular threshold θ_{thres} on the retina, where its upper and lower limit are $15^\circ \leq \theta_{thres} \leq 40^\circ$ respectively. T_C represents the time of collision between two objects, which can be expressed in seconds or the number of frames before the collision [165]. For the sake of simplicity, we use A_F representing the result of the detection. The mathematical description of A_F can be seen in Section 3.1.4.

Two metrics are used to evaluate the detection performance: the detection success rate (D_R) of the LGMD1 neuron and the false alarm rate (F_A) of the LGMD1 neuron. They are defined as follows.

$$D_R = \frac{\text{number of true detection}}{\text{number of image sequences}} \quad (3.29)$$

$$F_A = \frac{\text{number of false detection}}{\text{number of image sequences}} \quad (3.30)$$

3) Implementation: In previous research into LGMD1 and its postsynaptic target, the DCMD, the LGMD1 has been shown to respond selectively to approaching objects, either dark or light, but produces a weak response to translatory movements [33], [36], [62], [66]. Here, we have used a number of synthetic image sequences to test the basic characteristics of the model and have compared the performance of looming stimulus selectivity with two comparative models on synthetic and real datasets. The proposed model (TPM), the comparative model 1 (TCM1) [7], [31] and the comparative model 2

(TCM2) [8] are represented by symbols $*$, \circ and $+$ respectively in the graphical NMP curves. Note that the models of TCM1 and TCM2 have been introduced in Chapter 2 (see Fig. 2.6 and Fig. 2.7). The implementation of our experiments is as follows: we first presented and analyzed the different response properties between ON and OFF pathways by testing image sequences of looming and translational movement; we then verified the effectiveness of the denoising mechanism and the suppressing effect on wide-field movements; we finally compared the model with two baseline models on synthetic and real datasets.

3.2.2 ON/OFF Responses to Looming and Translating Stimuli

To investigate the response properties of moving edges to looming and translating stimuli, we first analyzed outputs of the P layer in time and spatial domains (see Fig. 3.6). Since edges of opposite contrasts (dark-to-light and light-to-dark) were processed in separate ON and OFF pathways [66], we have displayed the outputs from the dual channels and neural responses of the LGMD1 neuron (see Fig. 3.7).

As shown in Fig. 3.6, we recorded outputs of a specified pixel of the image plane in each frame (t axis) to obtain its responses in the time domain. Moreover, outputs of a horizontal line of pixels (x coordinate) in the spatial domain at a fixed time were also recorded. By comparing the curves of outputs in the time and spatial domains, We find that responses to looming and translating stimuli are quite different. The looming stimuli elicit OFF or ON single type outputs in t and x axis. However, the translating stimuli elicit both ON and OFF type outputs. We therefore infer that a comparison may exist between the ON and OFF channels inhibiting translating stimuli.

Fig. 3.7 further demonstrates such differences by presenting the neural responses of the ON and OFF pathways as well as of the neuron. As the neural response is related to the speed of movement and image contrast, we have tested the model with eight image sequences that are at constant, changing speed and opposite contrast. Fig.3.7(a) shows neural responses to looming motion. It can be seen that there is only one pathway in the model producing increasing excitations as the image size becomes larger. The LGMD1 neuron will produce a train of spikes after the expanding image size reaches a specific

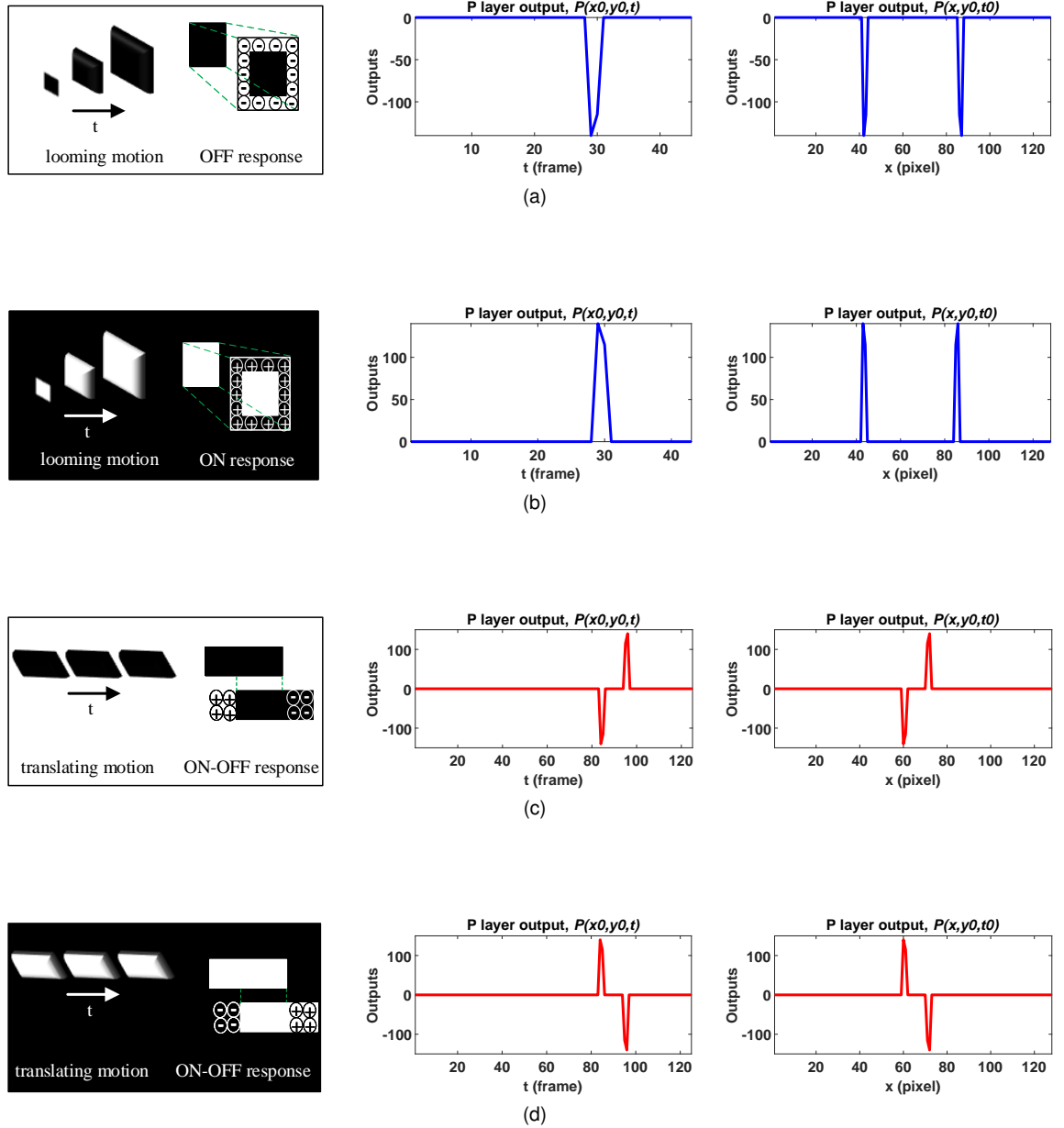


Figure 3.6: Outputs of the P layer when the model is tested with looming and translating stimuli. The left column presents looming or translational motion, including dark objects moving on light backgrounds as well as light objects moving on dark backgrounds. The middle column indicates outputs of the P layer of a specified pixel in the t axis. The right column denotes outputs of horizontal line pixels in the image plane at a fixed time point in the x axis. (a) and (b) show outputs of testing looming stimuli. (c) and (d) show outputs of testing translating stimuli.

size. Fig.3.7(b) displays neural responses to translational motion. In contrast, similar quantity excitations are produced in ON and OFF pathways to either the translating dark object or light object. As a result, the responses of LGMD1 are very weak or nearly close to zero.

According to Fig.3.7(a) and Fig.3.7(b), we can conclude that the response properties in ON and OFF pathways are quite different for looming and translational motion. Translatory moving objects produce paired ON-OFF responses while approaching objects yield a monotonous ON/OFF response. As the real biological LGMD1 neuron in the locust shows no or weak responses to translatory visual stimuli, it reveals a possibility of neural competition in ON and OFF channels in the visual systems of insects.

3.2.3 Effectiveness of the Denoising Mechanism

To enhance motion cues, a G layer that implements grouped excitation together with decay (GD) [7] processing is included after the S layer in the initial LGMD1 model. The G layer adopts a grouping strategy to gather together lateral excitation, thereby enhancing the response to coherent stimuli [67] while diminishing the small and isolated excitation produced by the background.

In this chapter, we have applied a denoising mechanism in the proposed model. The new mechanism has a similar mathematical form when compared to the GD processing strategy in the G layer. However, they are actually very different. Firstly, the structure of the LGMD1 model has been changed such that the functional layer that carries out the denoising is placed in advance of the S layer; secondly, the denoising mechanism of the D layer is realized by computing the expected pixel value within a 3×3 neighborhood so as to eliminate small and isolated excitations; finally, the original G layer needs an additional parameter to control the quantity of excitation, whereas the proposed model introduces a parameter of angle threshold to replace it.

To intuitively demonstrate the effectiveness of the denoising mechanism in the D_E/D_I layer, we tested the model with a set of image sequences with different levels of random noise. The noisy images produced by the computer can be mathematically

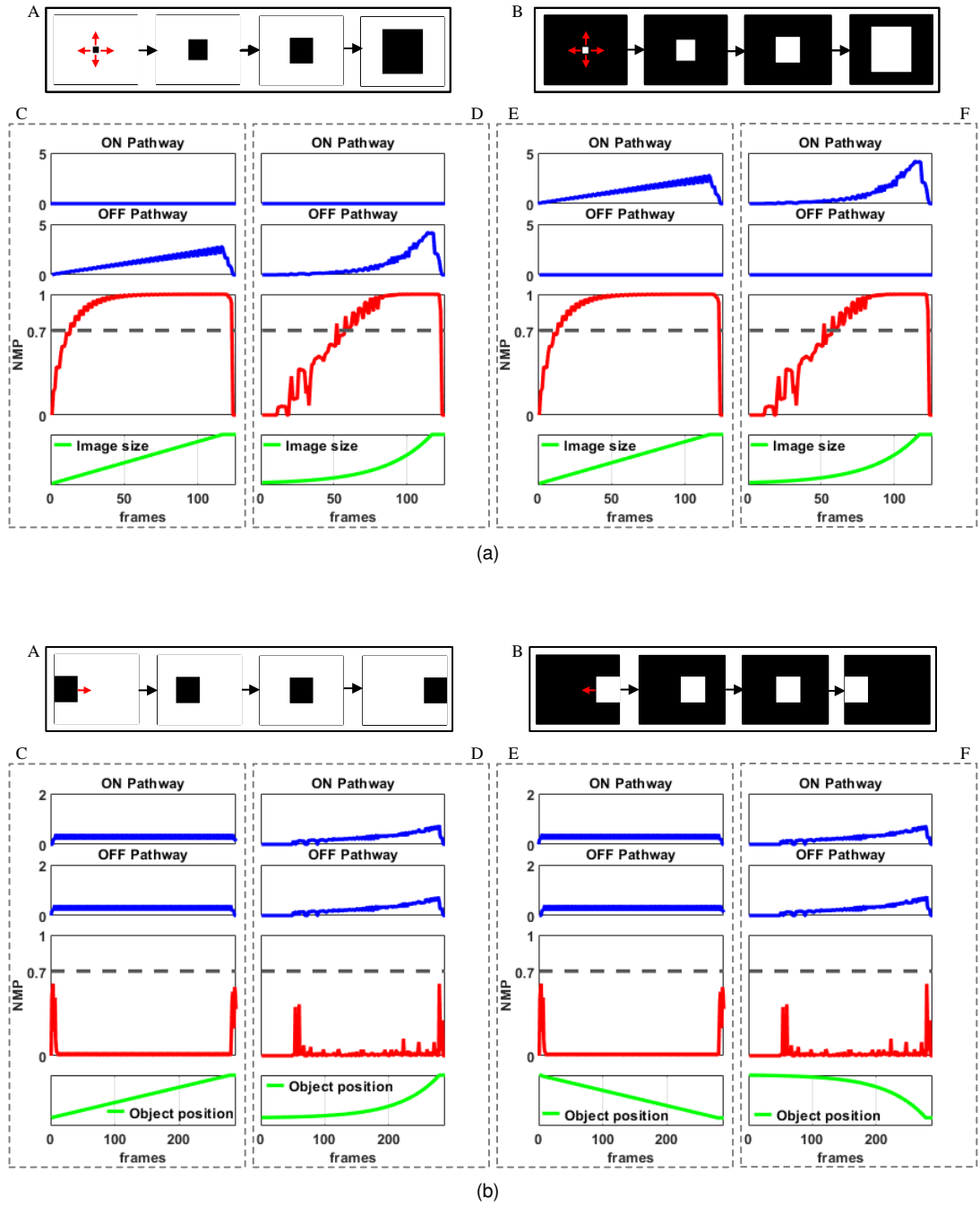


Figure 3.7: Responses to looming and translating stimuli respectively, including responding to ON and OFF channels, and the normalized membrane potential (NMP) of the LGMD1 neuron. The blue curve shows excitations from the ON and OFF pathways. The red curve denotes the normalized membrane potential (NMP) of LGMD1. The green curve represents the height of the square for looming stimuli and the position in the horizontal axis for translating stimuli independently. (a) A, B: approaching squares. C, E: response of looming stimuli at constant speed; D, F: responses to looming stimuli at loom-like increasing speed. (b) A, B: translationally moving squares. C, E: responses to translating stimuli at constant speed; D, F: responses to translating stimuli at loom-like increasing speed. Note that the spiking threshold is set at 0.7 in accordance with [8].

described as

$$\tilde{L}^n(x, y, t) = L(x, y, t) + k_n \cdot rand \quad (3.31)$$

where the function “rand” generates a single uniformly distributed random value between $(0, 1)$; k_n represents the level of random image noise and is set from $0 \sim 50$ units with a step increase of 10. Hence, there are six image sequences representing identical motions but with different noise levels. We have presented the denoising process (see Fig. 3.8) and have compared the neural responses under various processing strategies (see Fig. 3.9). Note that the parameter settings throughout the model without the G & D layer, with G layer and with D layer, are kept the same. This ensures that the differences produced by experimental results are inherent to the model structures.

Fig. 3.8 displays the procedure for extracting and denoising the motion signals. The input image is initially passed into the model and its random isolated excitations are inhibited by the denoising mechanism of the D_E/D_I layer. The exhibited output images of the E/I and the D_E/D_I layers in the separated ON/OFF pathways indicate random noise in the background can be effectively eliminated.

Fig. 3.9(a) presents sample frames of the looming image sequence to test the model. Fig. 3.9(b)-(d) displays neural responses of tested image sequences with various levels of noise. As can be seen, the outputs of the model with the G layer [9] or with the D layer are more consistent than without the G & D layers. This implies that the grouping excitation processing ([7], [67], [9]) or the denoising mechanisms are effective in diminishing random noise from the background. However, the model with the D layer results in an earlier alarm frame when compared to the model with the G layer. As the latency of the alarm frame may cause a failure in collision detection, we adopt the denoising mechanism (with the D layer) in the model.

3.2.4 Effectiveness of Suppression on Saccadic Eye Movements

Saccadic eye movements can be regarded as wide-field translatory movements. This is because when an animal moves its eyes or a robot turns its camera eyes, objects within the visual field seem to translate as their visual positions change, even though they have not moved [163]. In the visual system of the locust, the movement detec-

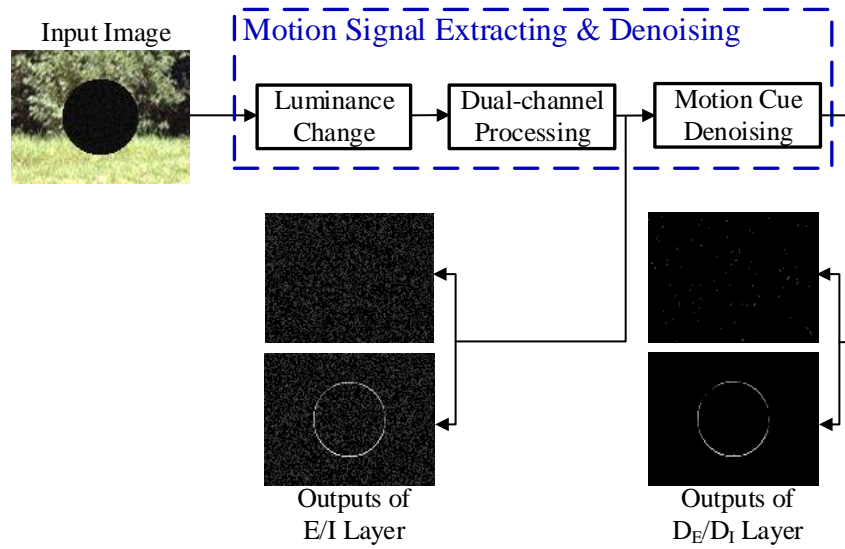


Figure 3.8: Schematic illustration of early visual processing in the proposed model. The input image was contaminated by random noise and it was passed into the P layer for extraction and denoising of motion stimuli. The processing stages were then implemented in the P, E/I, and D_E/D_I layers.

tor neurons are strongly inhibited during the fast phase of voluntary saccades [166]. Biological evidence suggests that the way that edges of opposite contrast inhibit each other may contribute to the reduction of responses to wide-field movements [66]. To verify the model's inhibitory effect on wide-field translational motion, we have used traditional sinusoidal gratings and moving backgrounds as the stimuli under test.

1) Tests Under Sinusoidal Gratings: The traditional stimuli created with sinusoidal gratings have been used to simulate the movement of visual scenes. Spatial and temporal frequencies are two main parameters which represent the motion of sinusoidal gratings. The model was challenged by a set of spatial frequencies (SF) and temporal frequencies (TF) individually. The example gratings, as input, are shown in Fig. 3.10(a). Fig. 3.10(b) shows the testing of sinusoidal gratings over a range of TF whilst maintaining the SF at 20 cycles. Fig. 3.10(c) shows the testing of sinusoidal gratings over a range of SF whilst maintaining the TF at 10 Hz. We found that the model did not respond to such stimuli even over a wide range of grating patterns. The lack of response to the sinusoidal gratings (Fig. 3.10(b), (c)) signifies that the paired ON-OFF responses are able to inhibit each other and the comparison between ON and OFF channels is effective. This demonstrates that the way the model responds to sinusoidal gratings is similar to that of a biological LGMD1 neuron.

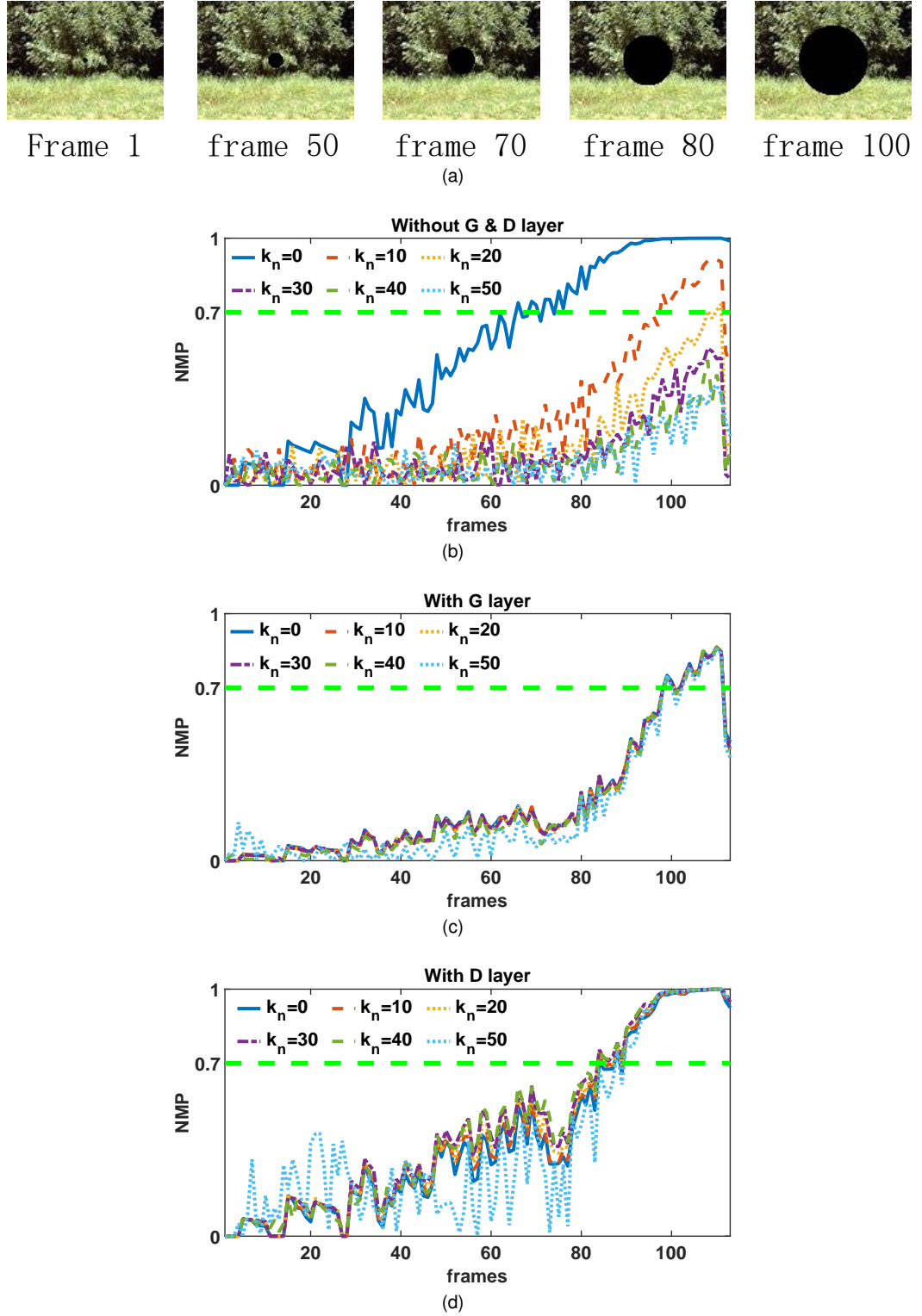


Figure 3.9: Neural responses of the proposed LGMD1 model that adopts various processing strategies. The visual stimuli tested consist of the image sequence with different levels of random noise ($k_n = 0, 10, 20, 30, 40, 50$). (a) The sample frames of the image sequence to test the model. (b) The model in the absence of either the GD processing or denoising mechanisms (without G & D layers). (c) The model implementing GD processing (with the G layer). (d) The model implementing the denoising mechanism (with the D layer).

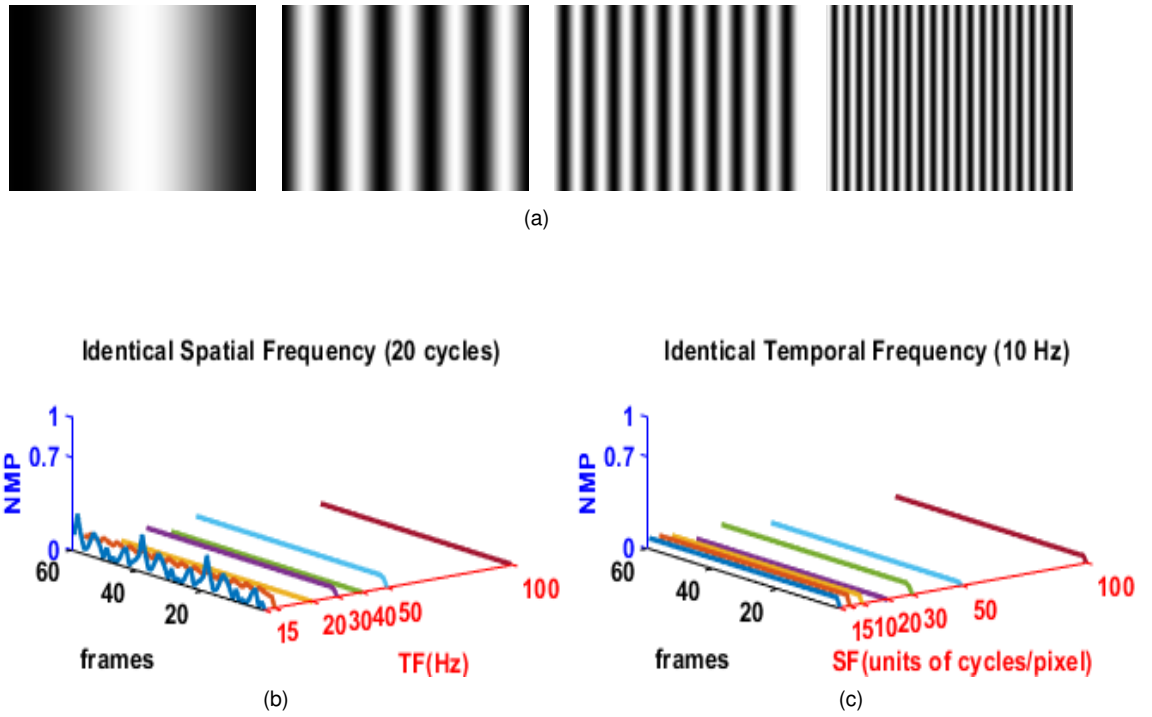


Figure 3.10: Neural responses to sinusoidal gratings with the spiking threshold set at **0.7**. (a) The example gratings as input. (b) The model is challenged by a series of gratings with different temporal frequencies. (c) The model is challenged by a series of gratings with different spatial frequencies.

2) Tests Under Background Movements: Considering the fast "saccadic" movements of animals [167] in the natural world, we have tested the model by subjecting it to changing natural scenes. The image sequences are separated into three groups (see Fig. 3.11(a)). Each group of background images have different textures reflecting different background types. The speed of movement of the natural scenes is 100 pixels per second and the sampling frequency is 100 frames per second. Neural responses of the model to three different kinds of moving backgrounds can be seen in Fig. 3.11(b). The results show that the translatory movements caused by moving backgrounds are suitably inhibited in most situations. Although a small number of unexpected spikes are produced by the third group of background images, they are triggered by the sudden appearance of objects in the visual field.

According to the above two sets of experiments, we have verified that the movement of backgrounds cannot evoke strong responses. This is because the numerous ON and OFF type stimuli created by changing visual scenes are sufficiently suppressed by the neural competition mechanism in the model. Hence, this competition mechanism makes the model insensitive to a wide range of translatory motions.

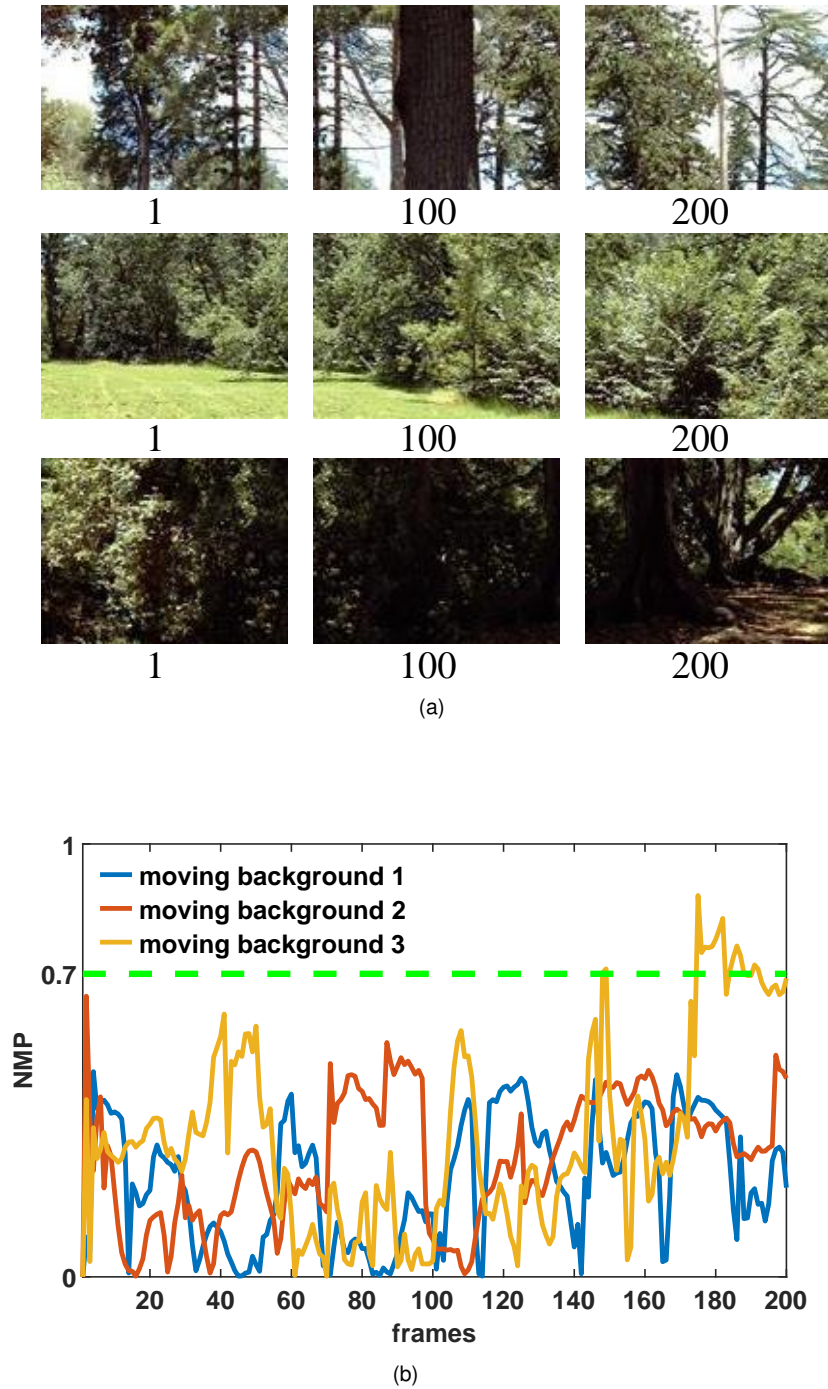


Figure 3.11: (a) Three video sequences of moving backgrounds are presented: top row (moving background 1), middle row (moving background 2), and bottom row (moving background 3). Each video sequence is presented with three frames; the frame number is indicated under each image. (b) Neural responses to three moving backgrounds.

3.2.5 Comparison on Synthetic and Real Datasets

In this section, the proposed model has been challenged by looming and translatory events. The proposed model was first tested on a synthetic dataset in terms of different object speeds and background luminance. Then, the real image sequences were utilized to further evaluate the performance of the proposed model. The performance comparison between the proposed model TPM and the two comparative models (TCM1 and TCM2), was also conducted. The three models are represented by symbols $*$, \circ , and $+$ in the normalized membrane potential (NMP) curves.

1) Comparison on synthetic dataset: The simulated motion of objects consists of a looming black disk ($1 \times 1 \sim 40 \times 40$ pixel), a 20×20 pixel black disk translating from left to right, and four 5×5 pixel black disks translating from the center to the left, right, up and down – i.e. four different directions. Sample images of these looming and loom-like translating events are presented in Fig. 3.12. The speed of the looming disk changed in an exponential manner. The single translating disk moved at changing speeds (loom-like), while the four translating disks moved in a divergent (loom-like) mode at constant speeds. The range of the speed was set to 10, 20, and 30 pixel/s. The object luminance $I_O = 0$ and the background luminance was set to four different values ($I_B = 1, 0.75, 0.5$ and 0.25). Hence the corresponding contrast $C = (I_O - I_B)/I_B$ is between the range of $(-1, -0.5)$.

Fig. 3.12 shows the NMP curves of the three models for visual stimuli at different speeds and contrasts. It can be seen that all models show good sensitivity to looming events. However, the loom-like translational motion at a fast speed or high contrast easily evokes spiking in the two baseline models, whereas the proposed model produces weak responses. The statistical results of these simulated video sequences are displayed in Table 3.2. From Fig. 3.12 and Table 3.2, we find that the proposed model works better than the two baseline models in inhibiting these simulated loom-like translatory events.

2) Comparison on real dataset: We further tested the developed model with recorded video sequences from real scenes, where the objects, e.g., balls, toy cars, and the camera, created looming and translatory events in an indoor environment with

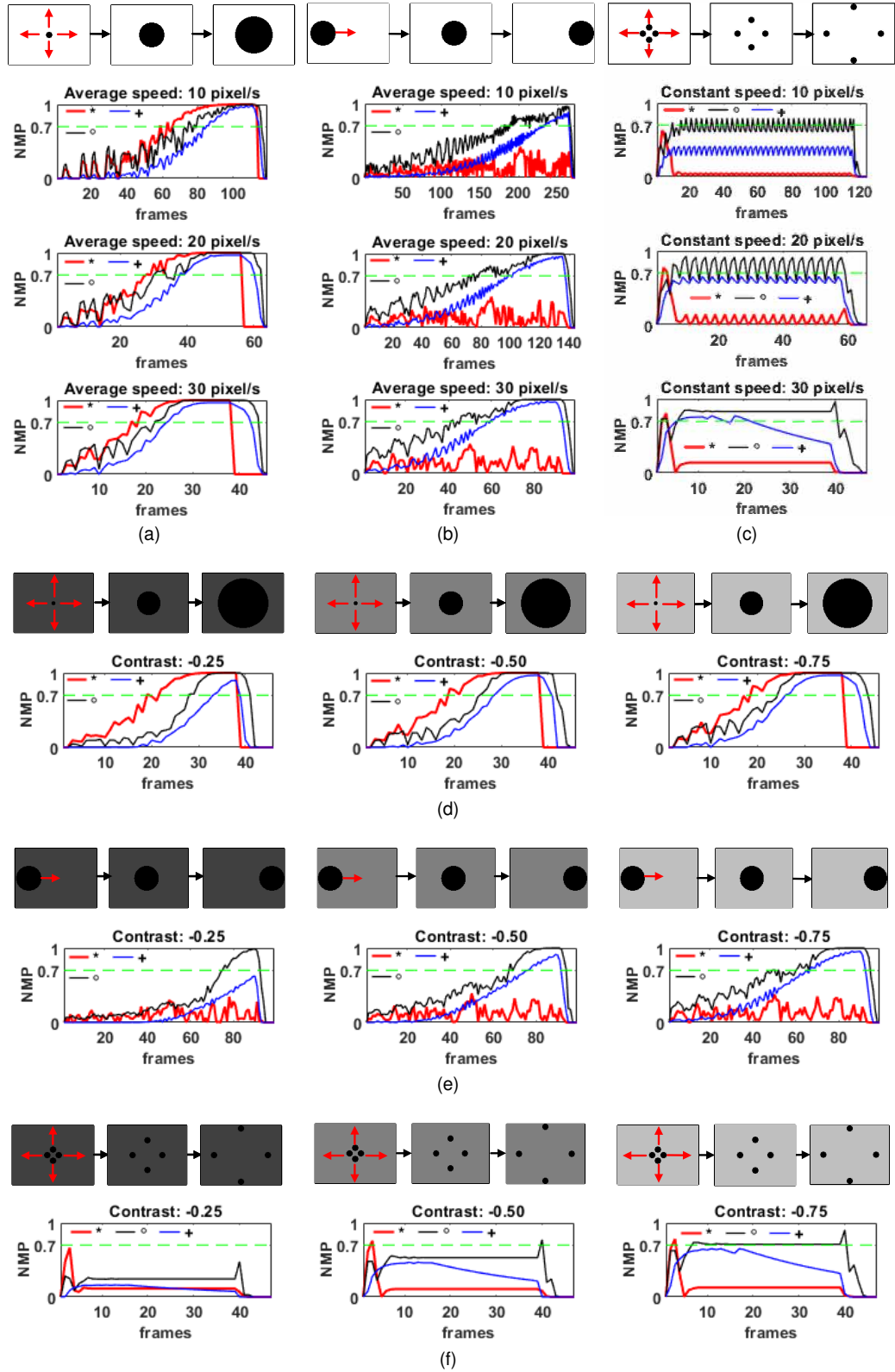


Figure 3.12: Example images of simulated motion events; neural responses of models to these moving stimuli; symbols *, o, and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. (a)-(c) Output curves of looming versus translational events at different speeds. (d)-(e) Output curves of looming versus translational events under different contrasts. **The red curves show that the proposed model responds selectively to looming objects at various speeds/contrasts but not to translating objects, while other models show no preference between the looming and the translating objects.**

Table 3.2: Detection success rate (D_R) and false alarm rate (F_A) for the synthetic visual stimuli experiments. (see Fig. 3.12).

	TPM		TCM1		TCM2	
	D_R	F_A	D_R	F_A	D_R	F_A
Looming events	1	0	1	0	1	0
Translating events	N/A	0	N/A	0.67	N/A	0.5

TPM: The proposed model

TCM1: The comparative model 1

TCM2: The comparative model 2

simple and complex backgrounds. Fig. 3.13 presents the performance of a single ball approaching the camera and translating from left to right against simple backgrounds as well as two translatory balls moving from the center to left and right separately. Similar motion events with complex backgrounds can be seen in Fig. 3.14, but the moving objects are toy cars. Fig. 3.15 displays the movements of the GoPro camera, involving the approach of toy cars or turning in front of them. Note that the turning of the camera is regarded as the translatory motion of a wide field of view.

From the neural response curves of the LGMD1 models in Fig. 3.13 - Fig. 3.15, we can clearly see that the two comparative baseline models produce false alarming frames so that they cannot differentiate looming motion from translational motion. As the proposed model only responds vigorously to looming events, it is robust in suppressing the translational events. Table 3.3 exhibits the statistical results of these real video sequences. It also demonstrates that the proposed model works very well in detecting looming cues against the translational motion.

Table 3.3: Detection success rate (D_R) and false alarm rate (F_A) for the real visual stimuli experiments. (see Fig. 3.13-3.15).

	TPM		TCM1		TCM2	
	D_R	F_A	D_R	F_A	D_R	F_A
Looming events	1	0	1	0	1	0
Translating events	N/A	0	N/A	1	N/A	1

TPM: The proposed model

TCM1: The comparative model 1

TCM2: The comparative model 2

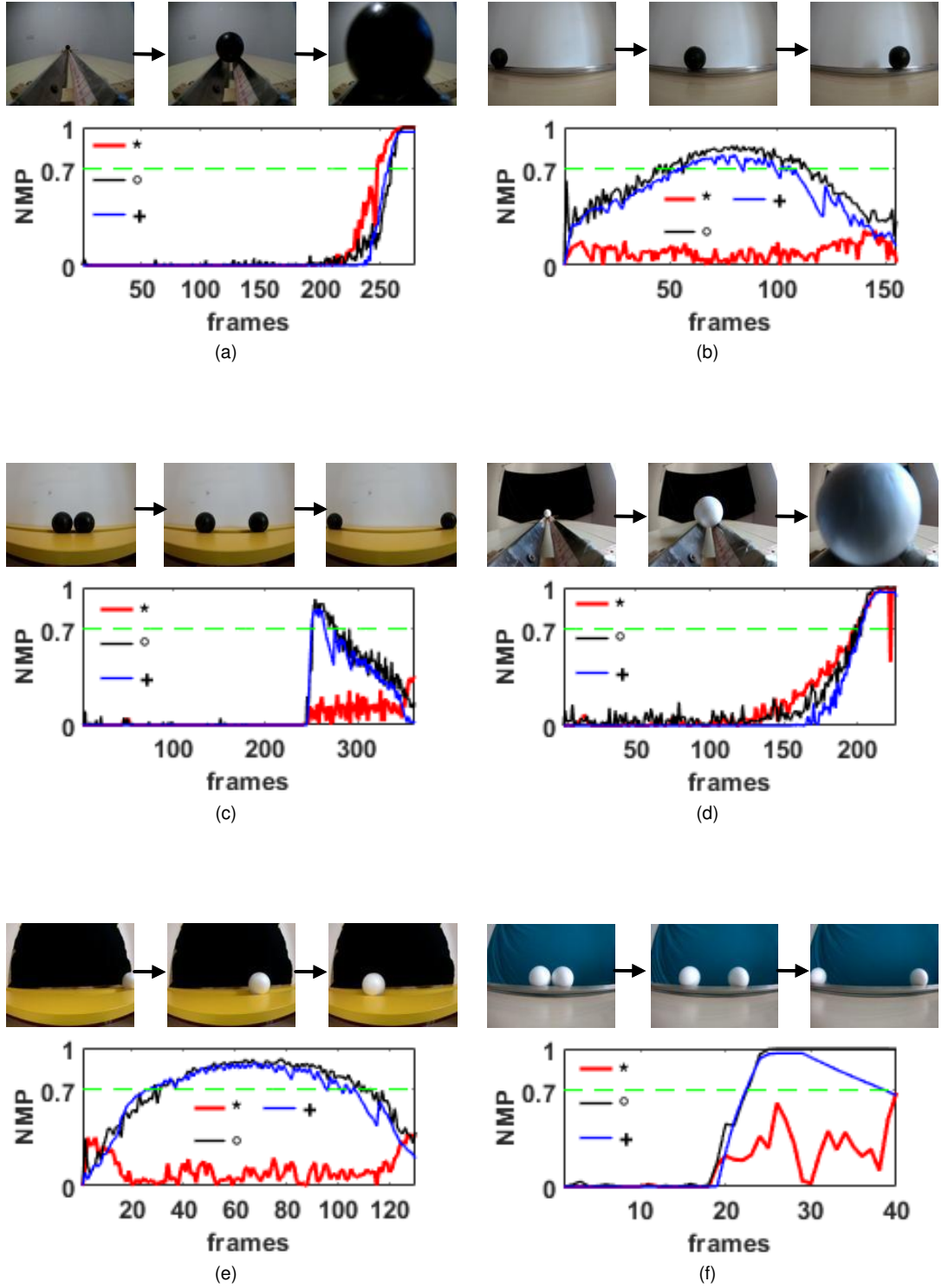


Figure 3.13: Example images of real motion events with simple backgrounds; neural responses of models to movements by objects; symbols *, \circ , and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. The object approaches the camera: (a) the black ball, (d) the white ball. A translating object moves from left to right in front of the camera: (b) the black ball, (e) the white ball. Two translating objects move from the center to left and right separately: (c) two black balls, (f) two white balls. **The red curves show the proposed model prefers to respond to looming objects rather than translating ones against clean backgrounds.**

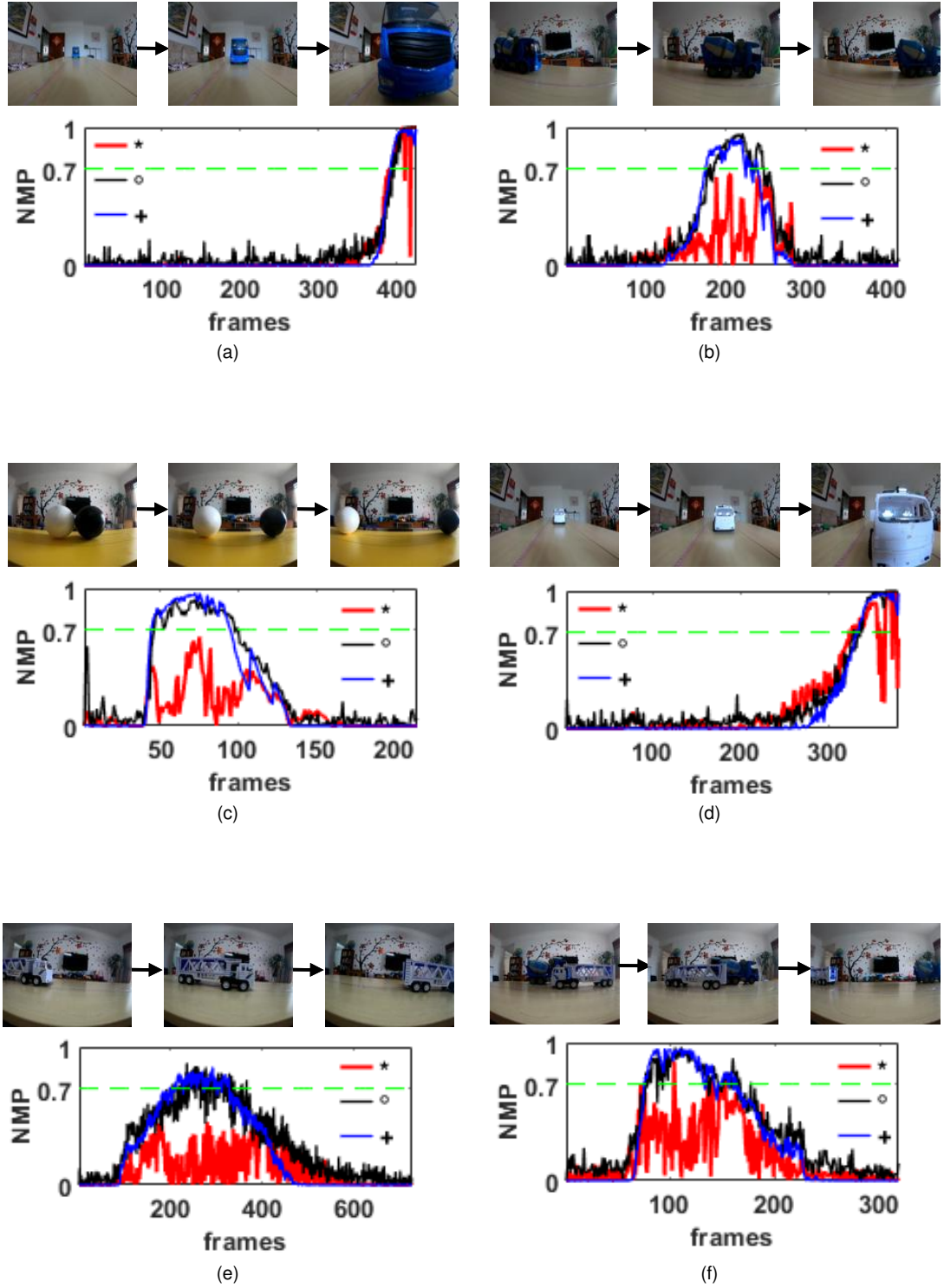


Figure 3.14: Example images of real motion events with complex backgrounds; neural responses of models to movements by objects; symbols *, \circ , and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. The object approaches the camera: (a) the blue toy car, (d) the white toy car. A translating object moves from left to right in front of the camera: (b) the black ball, (e) the white ball. Two translating objects move from the center to left and right separately: (c) the black ball and the white ball, (f) the blue toy car and the white toy car. **The red curves show the proposed model prefers to respond to looming objects rather than to translating ones against complex backgrounds.**

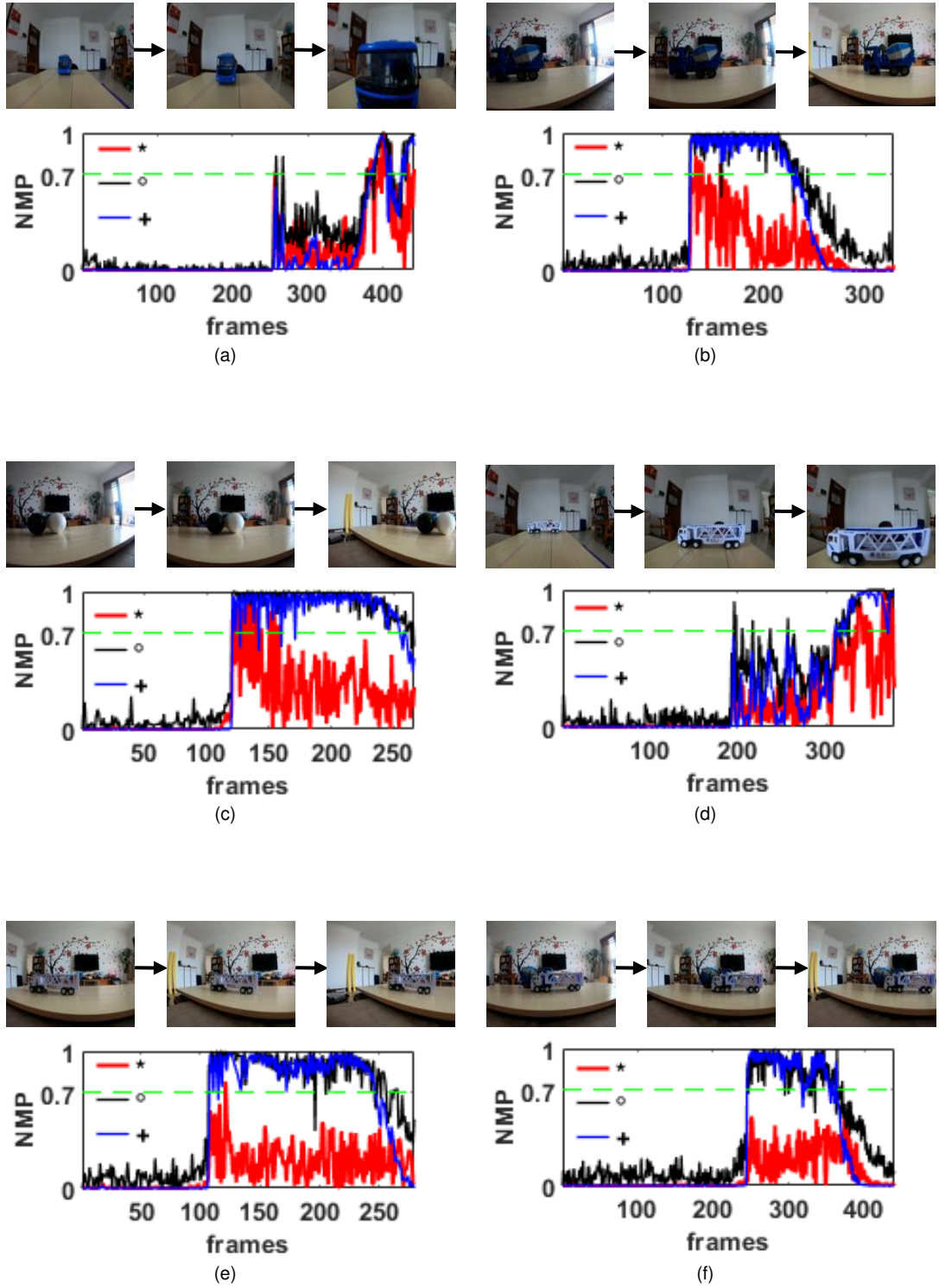


Figure 3.15: Example images of real motion events with complex backgrounds; neural responses of models to movements by the camera; symbols *, o, and + represent the proposed model, the comparative model TCM1 and the comparative model TCM2 respectively. The camera approaches the object: (a) the blue toy car, (d) the white toy car. The camera turns from right to left in front of objects: (b) the blue toy car, (e) the white car, (c) the black ball and the white ball, (f) the blue toy car and the white toy car. **The red curves show the proposed model prefers to respond to looming objects rather than other translating cues with the camera in motion.**

3.2.6 Comparison with Conventional Methods

Conventional obstacle detection methods can be categorized into optical-flow based [90], [91] and feature based [95], [96] methods. The former utilizes the time to collision (TTC) information from the optical flow to detect obstacles, including two methods. One is the OF-E method [90] which is relying on the ratio of distance against the flow speed. The distance here can be measured from the point on the optic flow image plane to the focus of expansion. Another is the computation of flow divergence (OF-D) [91] method. The latter, the feature based methods [95], [96], utilizes the expansion ratio of relative feature sizes (FD-S) [95] and area (FD-A) [96] to detect obstacles.

We have compared the proposed LGMD1 model (TPM) with the four types of classic and state-of-the-art collision detection methods, i.e., OF-E [90], OF-D [91], FD-S [95] and FD-A [96]. The accuracy (Acc) is employed to quantitatively evaluate the models' performance, which can be computed as [26]:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.32)$$

where TP (true positive), TN (true negative), FP (false positive), and FN (false negative) refer to categories of the outputs. The higher the accuracy, the more accurately the looming object can be detected. TTC is the output of OF-E and OF-D. The ratio is the output of FD-S and FD-A. Normalized membrane potential (NMP) is the output of TPM. Note that a collision is defined to occur if the value of $TTC \in (0, 0.04s)$ at the frame (frame-to-collision). In the experiments, we report the maximum accuracy for each method.

The proposed LGMD1 model as well as the other four conventional methods are evaluated on the Vision Egg dataset [164], where sample images are shown in Figs. 3.16-3.19. The frame rate of synthetic sequences is 30 fps. These image sequences tested are categorized into black disks against the white background, textured objects against the white background, black disks against the complex background and textured objects against the complex background. Each group involves looming and trans-

lating events, e.g., an object looming at changing speed or constant speed, an object translating at changing (loom-like) speed, and four objects translating at constant speed in a divergent (loom-like) mode.

The OF-E method uses Horn and Schunck (HS) [83] technique to estimate optical flow. The focus of expansion point is set at the center of the image plane simply before computing the distance since the looming object is on a head-on collision trajectory. The OF-D method applies FAST algorithm [93] to detect corners before using the Lucas-Kanade (LK) tracker [89]. The TTC is reversely related to the divergence [92] that is computed by averaging ratios of image distances between the every two tracked corners in the OF-D method. The FD-S method uses template matching to confirm the expansion ratio of sizes after detecting SIFT/SURF features [99], while the FD-A method compares the ratio of area based on the detected feature points.

The comparison of the computational cost of processing the four groups of image sequences among the five methods is shown in Table 3.4. Note that the computational cost of the OF-D method is low when the numbers of detected corners are very small or none as LK is carried out only at these corner points. It is also worth mentioning that the HS, LK, FAST and SURF algorithms adopted in four conventional methods are functions in the Matlab Toolbox that are optimized. As can be seen, the proposed LGMD1 model needs to be in the region of 0.0036s to process a frame, which is more efficient than the other four obstacle detection methods.

Table 3.4: Running time comparison of the five methods on the four groups of image sequences in vision egg datasets. The value is averaged over 1370 frames with a frame size of 128 pixels (horizontal) and 96 pixels (vertical).

	OF-E	OF-D	FD-S	FD-A	TPM
Time (s/frame)	0.0156	0.0071	0.0165	0.0112	0.0036

For performance comparison, the outputs of various methods in the four groups are provided in Figs. 3.16-3.19 separately. As can be seen from Figs. 3.16-3.19, the proposed LGMD1 model can **detect looming objects accurately under conditions that are lacking texture, moving at changing speed, or against complex backgrounds.** However, the other four methods are unable to detect obstacles accurately, or produce

false detection results when translating objects move in a loom-like mode. For qualitative evaluation, we report an accuracy comparison in Table 3.5. The accuracy of the proposed LGMD1 model is higher than the previous best performing method by 0.5, 0, 0.25 and 0.25 in terms of accuracy on the four groups of image sequences.

The looming motion detectors should work timely and reliably in a wide range of conditions since the real visual world is complex and dynamic. The results mentioned above demonstrated the robustness and efficiency of the proposed method compared to the other four conventional methods under different conditions.

Table 3.5: Accuracy comparison of the five methods on the four groups of image sequences. The larger the accuracy, the more accurately the looming object can be detected.

	OF-E	OF-D	FD-S	FD-A	TPM
Group 1	0.50	0.25	0.50	0.50	1.00
Group 2	0.50	1.00	0.75	1.00	1.00
Group 3	0.50	0.75	0.75	0.75	1.00
Group 4	0.50	0.75	0.50	0.75	1.00

3.3 Chapter Summary

In this chapter, we have proposed a new bio-plausible LGMD1 model based on the comparison of responses from ON and OFF pathways. The computational architecture of the model consists of six layers which integrate neural information-processing mechanisms for extracting cues for looming motion. To intuitively illustrate the probability of neural competition and the effectiveness of the denoising layer, we present different ON/OFF responses of looming and translational motion separately as well as exhibit the performance of the model when applying the denoising mechanism. Moreover, tests of the suppression effect on moving backgrounds indicate that neural competition is also effective in suppressing responses to wide-field motion. Finally, comprehensive experimental comparisons with other models demonstrate that the proposed neural network is able to robustly inhibit a response to any translational motion within the field of view. The results demonstrate that the model is capable of

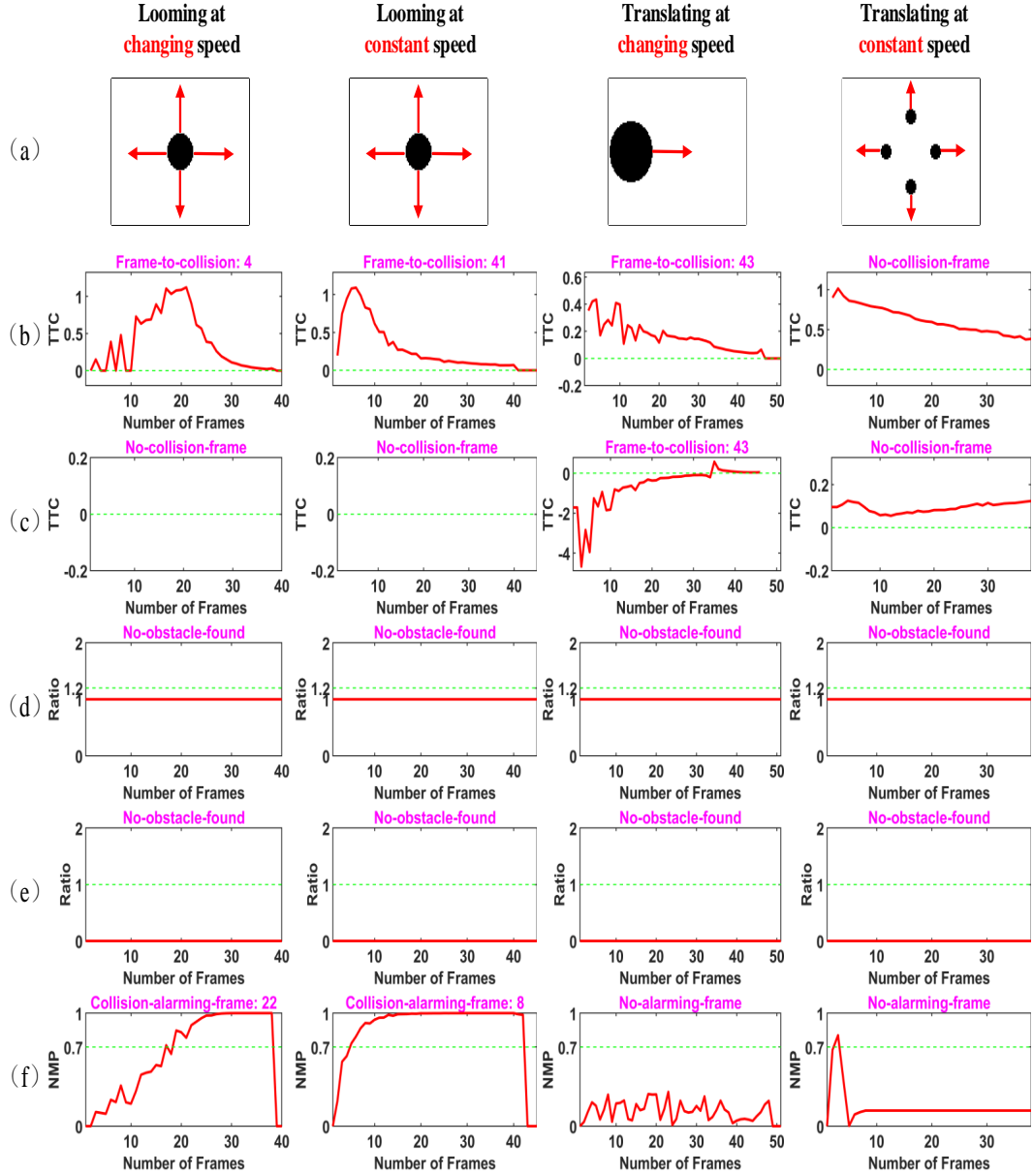


Figure 3.16: Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 1** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E, which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into NP, NP, NP and TN from column one to column four. (d) The outputs of FD-S, which are categorized into NP, NP, TN and TN from column one to column four. (e) The outputs of FD-A, which are categorized into NP, NP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that untextured object against white background.**

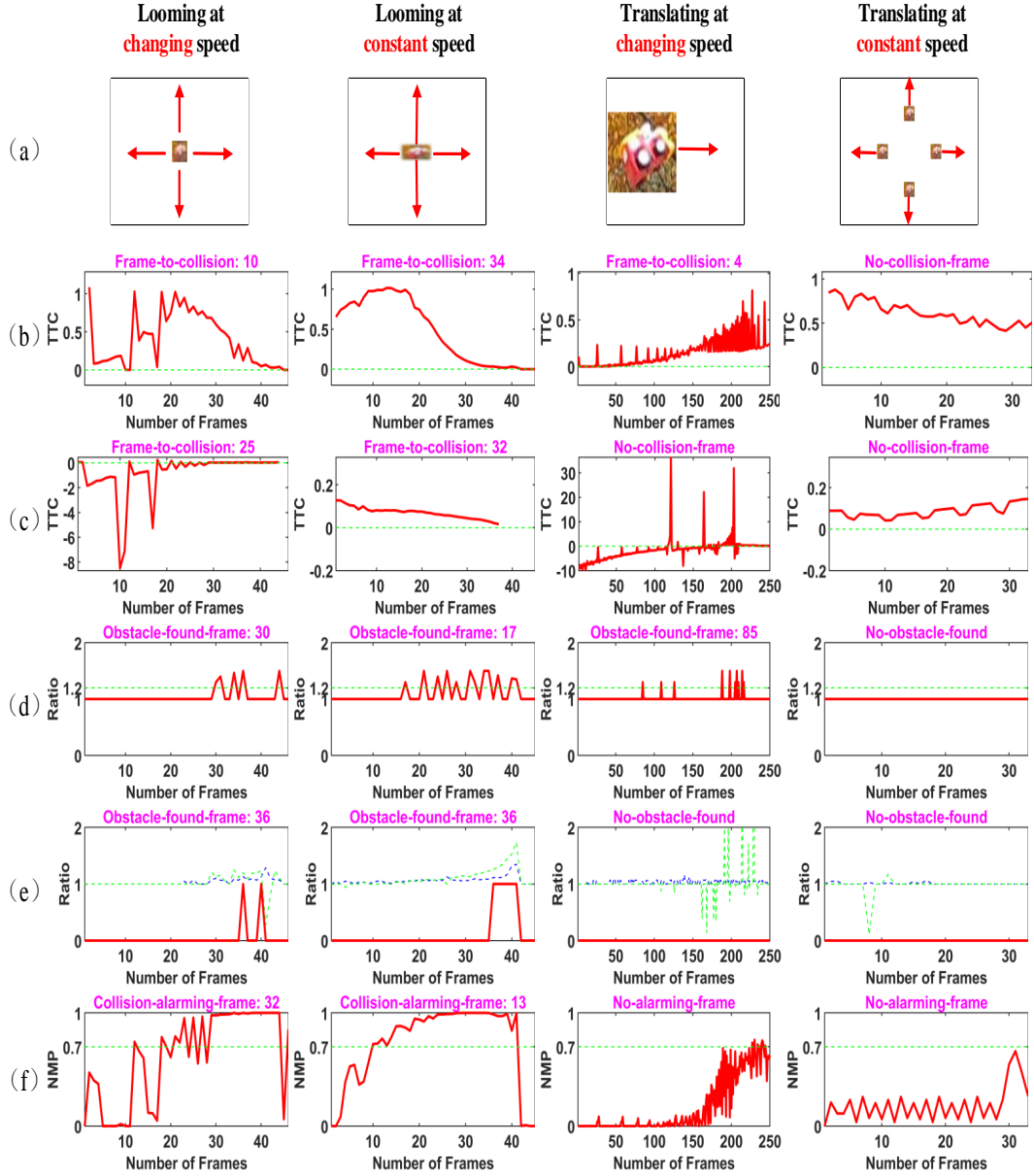


Figure 3.17: Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 2** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E, which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into TP, TP, TN and TN from column one to column four. (d) The outputs of FD-S, which are categorized into TP, TP, FP and TN from column one to column four. (e) The outputs of FD-A, which are categorized into TP, TP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that textured object against white background.**

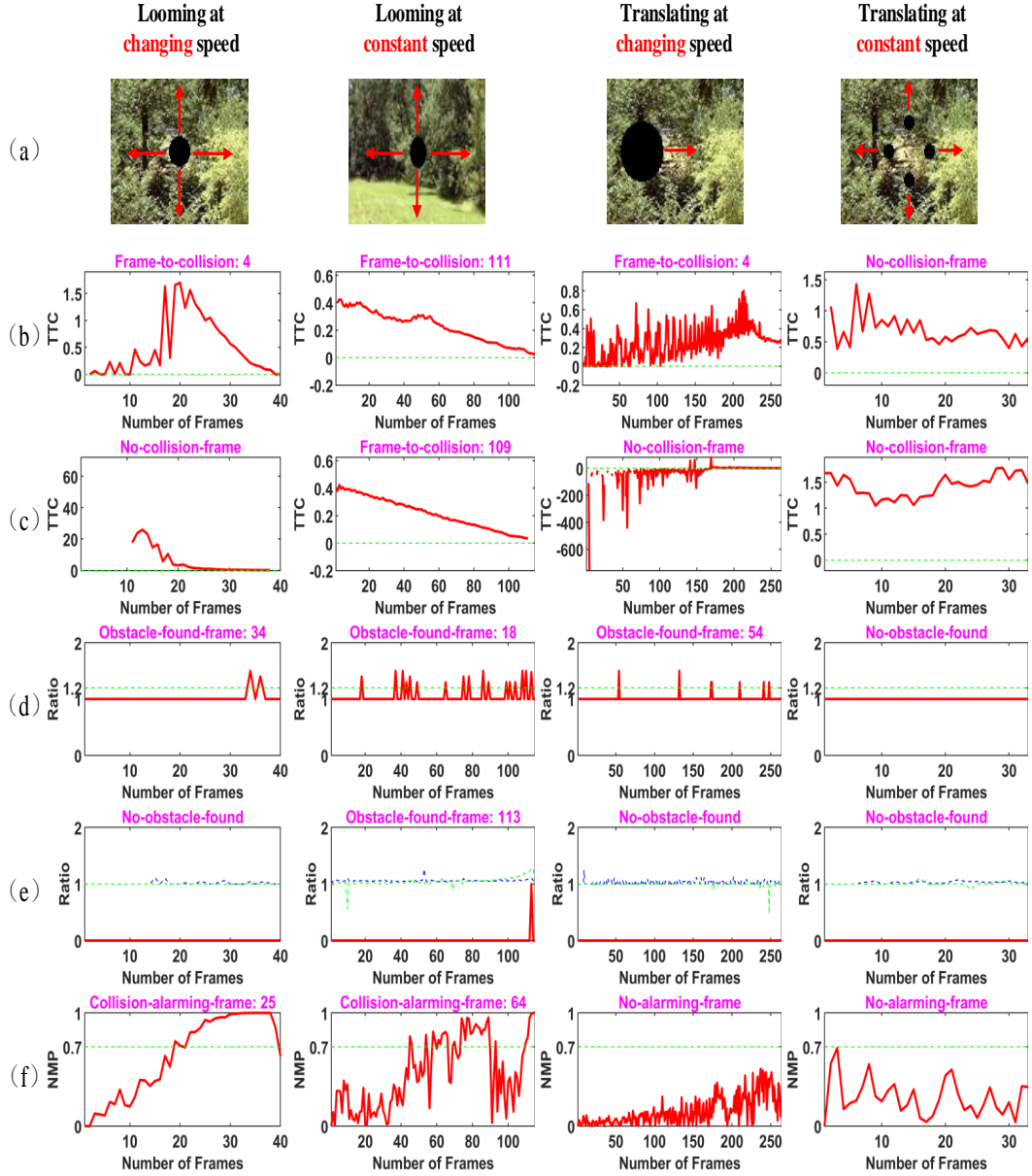


Figure 3.18: Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 3** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E, which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into FN, TP, TN and TN from column one to column four. (d) The outputs of FD-S, which are categorized into TP, TP, FP and TN from column one to column four. (e) The outputs of FD-A, which are categorized into FN, TP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that untextured object against complex background.**

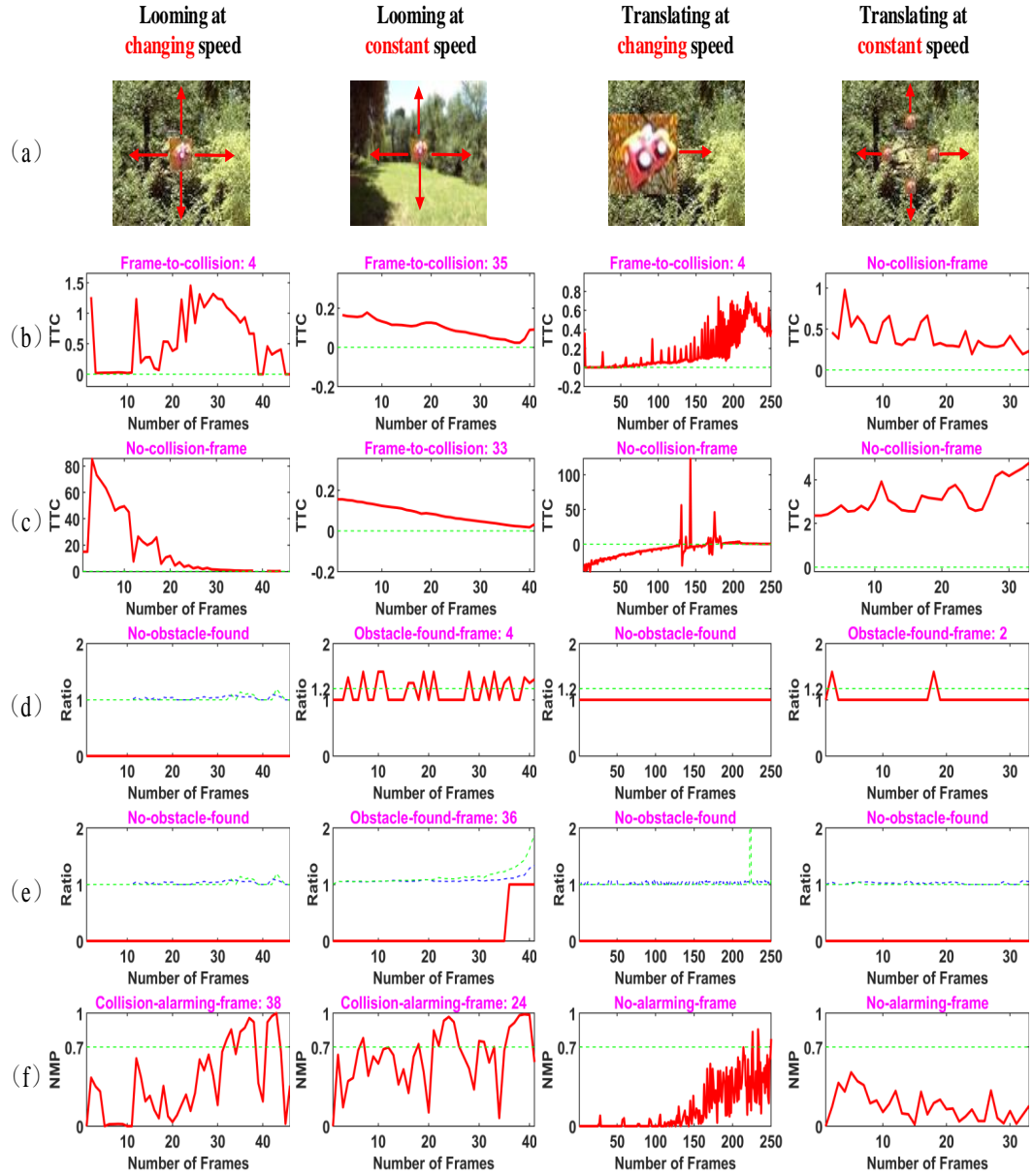


Figure 3.19: Comparisons between the proposed LGMD1 model and other four conventional methods (rows) when they are tested by four image sequences (columns). Four sample images from the four image sequences in **Group 4** (row a), and outputs of various methods (row b-f). (a) Input image. (b) The outputs of OF-E, which are categorized into FP, TP, FP and TN from column one to column four. (c) The outputs of OF-D, which are categorized into FN, TP, TN and TN from column one to column four. (d) The outputs of FD-S, which are categorized into FN, TP, TN and FP from column one to column four. (e) The outputs of FD-A, which are categorized into FN, TP, TN and TN from column one to column four. (f) The outputs of TPM, which are categorized into TP, TP, TN and TN from column one to column four. **Compared with other four methods, the proposed LGMD1 model is robust in detecting looming motion and does not produce false collision alarms under the condition that textured object against complex background.**

responding correctly to approaching and translatory movements, enhancing collision selectivity.

In the future, we will further investigate the potential applications of the proposed looming-sensitive model, which aims to handle more complex and dynamic visual scenes for the navigation of intelligent robots and vehicles.

Chapter 4

Dark Adaptation- Low Light Image Enhancement

The enhancement of low-light images is essential for vision-based autonomous robots navigating under low illumination [168]. This is because the low-light images perceived by the robots are usually with low intensity and low contrast, which brings challenges for visual tasks like motion detection. Therefore, enhancing the low-light image becomes very important for detecting faint movements in low light conditions. Moreover, motion detection relies on extracting luminance changes [9], [142], and image enhancement methods should not introduce too many unwanted artifacts or should prevent serious lightness or contrast distortion to protect the original motion information. This indicates that the preservation of the naturalness [135] of the image is important for enhancement methods.

For enhancement processing of low-light images, intensity transformation and image denoising are two highlighted aspects. The former aims to implement nonlinear operations to achieve raising the intensity of the dark pixel. The latter emphasizes the reduction of noise produced by the intensity amplification processing. Specifically, intensity amplification is a crucial step in disclosing details buried in dark regions. Here, we focus on adaptively raising intensities to overcome the problem of overenhancement and preserve luminance naturalness.

Unlike human eyes, nocturnal insects' eyes still possess remarkable visual abilities even though illumination levels are extremely low [11]. The dark adaptation in insects'

visual systems can explain why they can see color and detect faint movements in very dim environments [10]. For example, the amplitude of the photoreceptor responses (“bumps”) in the nocturnal species is about five times the amplitude of those in the day-active species due to higher phototransduction and membrane gains [116], [169]. Hence, the underlying neural mechanisms of dark adaptation may provide us with useful solutions to adaptively raise intensities.

Although there are various types of low-light image enhancement methods [20], [21], [170]–[172], such as the histogram-based method [170], the Retinex based method [20], the gradient-based method [21], etc, these methods mainly rely on the statistical information of intensities, the estimation of illumination, or other fusion information like the gradient. However, they do not consider the ever-changing illumination situations, especially for dim-light scenes with motions, which is required to raise the intensity adaptively as well as reduce the distortion.

For dark adaptations, biological researches are devoted to building plausible models to interpret the results of electrophysiological experiments, which can better explain the relationship between the captured photons and the response amplitude. Correspondingly, a lot of models were proposed for explaining the adaptations of receptors in different aspects, including Stevens’ psychophysical power law adaptation mechanism [108], [173], the shunting synaptic inhibition mechanism [109], [174] and canonical neural computations [15], [111], etc. The phototransduction process is very complex, and the complete mechanisms underlying photoreceptors’ dark adaptation remain unclear. Nonetheless, the exploration of nocturnal vision mechanisms is always a promising way to develop effective and efficient image enhancement algorithms.

In this chapter, we propose a bio-inspired low-light image enhancement method, involving the dark adaptation processing in R, G and B channels. The dark adaptation is composed of a series of canonical neural computations, and the adaptation parameter in each color channel is related to the different wavelength photoreceptor cell. The proposed bio-inspired adaptation method can not only adaptively raise intensities but also preserve the naturalness. The main contributions of this work are: 1) a dark adaptation based framework is proposed for low-light image enhancement, which consists

of a series of canonical neural computations; 2) combining the psychophysical power law with the photoreceptor's response properties to different wavelengths of light during the application of dark adaptation for enhancing color image; 3) comprehensive analysis of parameters of the model and quantitative comparison of performance on image quality.

The remainder of this chapter is organized as follows. Section 4.1 introduces the proposed bio-inspired dark adaptation framework. Section 4.2 reports comprehensive experimental results as well as performance comparisons with existing methods on the low light image dataset. Section 4.3 presents the discussion. Finally, we conclude this chapter in Section 4.4.

4.1 Bio-inspired Dark Adaptation Framework

In this section, we present the proposed bio-inspired dark adaptation framework. The key idea of the dark adaptation is to adaptively raise the intensities of dark pixels by a series of canonical neural computations. Given an input image, it is first sampled into three channels, that is, converted to the red (R), green (G), and blue (B) channel images. As shown in Fig. 4.1, the input image implements the dark adaptation in three separate channels to enhance the color components. It is worthy to note that

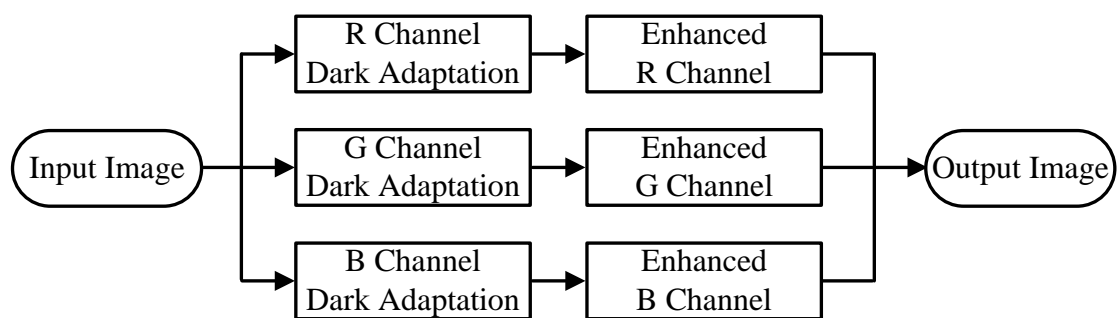


Figure 4.1: Proposed dark adaptation framework for low light image enhancement. The red (R), green (G), and blue (B) components of the input image are processed with the dark adaptation in three separate channels. Note that each channel has a different adaptation parameter.

the dark adaptation processing within the three color channels has different adaptation parameters. This is because the three color channels correspond to the three types of photoreceptor cells which are sensitive to light with different wavelengths. Finally, the

final output image can be obtained by combining the enhanced images from the three color channels.

The proposed dark adaptation in this chapter consists of a series of canonical neural computations, as illustrated in Fig. 4.2. It involves the power law adaptation, di-

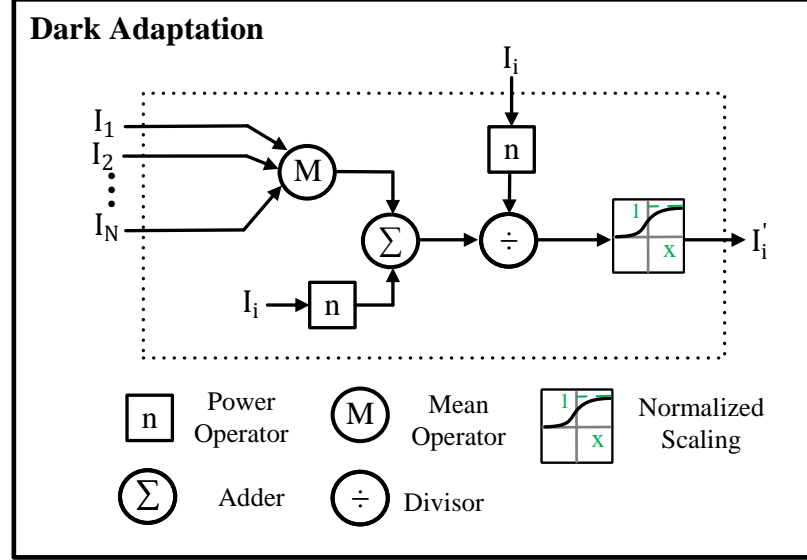


Figure 4.2: Schematic illustration of the proposed dark adaptation. There are N cells that correspond to N pixels in the input image, denoted by $I_1 \sim I_N$. n denotes the sensation parameter, and its value depends on the wavelength of perceived light. I_i and I'_i indicate the i th cell and its enhanced output after the dark adaptation processing. For clear illustration, we only give one cell's enhanced result.

visive normalization and adaptive rescaling operations. First, the power law operation compresses the input intensity into a narrow range. Then, the divisive normalization adaptively maps the power law output into a range of $0 \sim 1$ controlled by the average intensity of the input image. Finally, the normalized output is remapped by adaptive rescaling operation, which can guarantee the maximum transmission of information. In the following subsections, more details about these neural computations mentioned above are introduced as follows.

4.1.1 Power Law Adaptation

For power law adaptation, the famous Stevens' psychophysical power law is deduced mathematically from the Weber-Fechner logarithmic law [108]. Moreover, many researchers claimed that the power law system could adjust its effective adaptation timescale to the environment [175]. The functionality of the power law adaptation is

able to compress the physical stimuli into a specific range of perceived intensities by the different parameters of power. The classical power law equation is given by [108],

$$\psi = a \cdot (I - I_0)^n \quad (4.1)$$

where ψ is the perceived intensity, I is the physical intensity, and a , I_0 and n are constants. Note that the exponent n is related to the sensory system. Here, the value of n is related to the color channel. This power law equation describes relations between sensation magnitude and stimulus strength ranging from zero to increased strength with an upper limit of 255. We set $a = 1$ and $I_0 = 0$, then Eq. (4.1) has the simplest form $\psi = I^n$. The response amplitude is correspondingly only determined by the adaptation parameter n . In many invertebrate photoreceptors, the intensity/response functions are generally fitted with n less than one [12]. Hence, we display the perceived intensity results resulting from the adaptation parameter n in the range of $0.1 \sim 0.9$ (see Fig. 4.3). During the processing of dark adaptation, the perceived intensities in three color

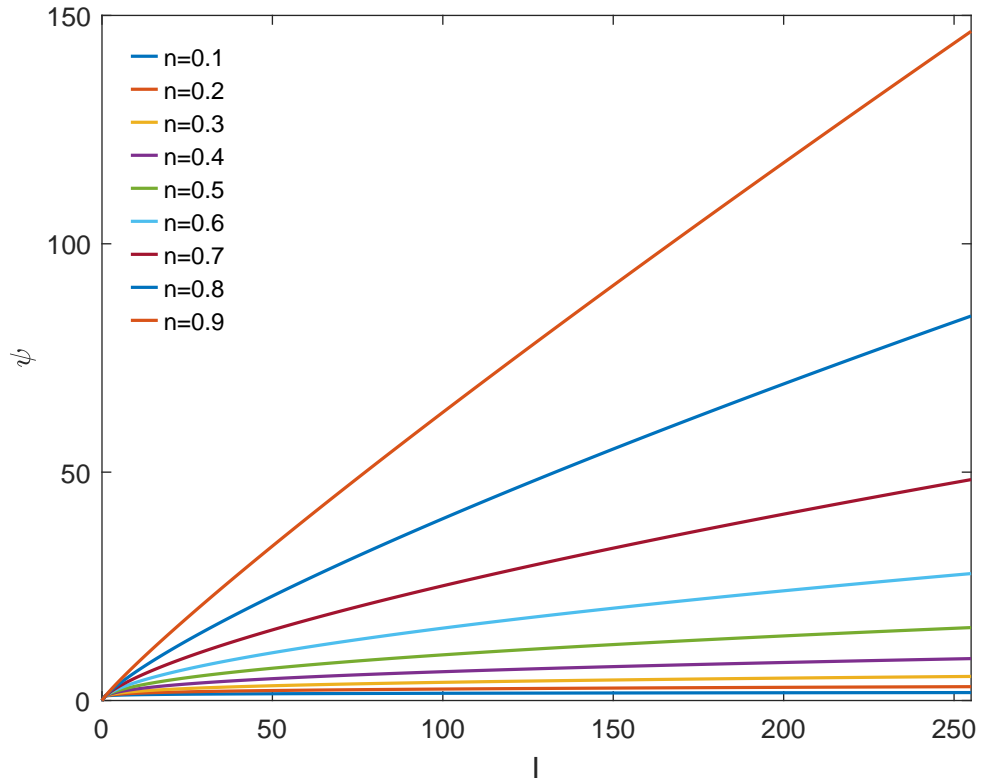


Figure 4.3: The perceived intensity ψ for different values of the adaptation parameter n (refer to Eq. (4.1), $a = 1$, $I_0 = 0$).

channels are given by,

$$\psi_R = I_R^{n_r} \quad (4.2a)$$

$$\psi_G = I_G^{n_g} \quad (4.2b)$$

$$\psi_B = I_B^{n_b} \quad (4.2c)$$

where the value of n_r , n_g and n_b relates with the sensation parameters of long-, medium-, and short- wavelength photoreceptor cells, responding to red light (620–750 nm), green light (495–570 nm) and blue light (450–475 nm). To determine the suitable parameters, we calculate the root-mean-square (RMS) level of the absolute difference image between the output image (power law operation for dark image) and the reference image since the RMS can be used to estimate the transformation biases [176]. We choose four groups of parameters to calculate the RMS. Each group has six different combinations of n_r , n_g and n_b , and the minimum value of RMS corresponds to the optimal combination. More details on choosing the adaptation parameters n_r , n_g and n_b are presented as follows.

Fig. 4.4 shows the reference image and the dark image. First, the reference image



(a)



(b)

Figure 4.4: The same scenes are captured by a GoPro camera under different illumination conditions. (a) Reference image (normal illumination). (b) Dark image (low illumination).

$R(x, y)$ and the dark image $I(x, y)$ are converted into R, G and B channel images separately (i.e., $R_c(x, y)$, $I_c(x, y)$, $c \in (R, G, B)$). Then, the color components of dark image are transformed by the power law operation (see Eq. (4.2a-4.2c)) producing

new images $\psi_c(x, y)$. Finally, the average RMS ($\overline{\text{RMS}}$) on the three color channels is computed by

$$\overline{\text{RMS}} = \frac{1}{3} \sum_{c=1}^3 \sqrt{\frac{1}{M * N} \sum_{x=1}^M \sum_{y=1}^N |\psi_c(x, y) - R_c(x, y)|^2} \quad (4.3)$$

where M and N denote the number of rows and columns of the input image. As the fitted value of the adaptation parameter in some invertebrate photoreceptors is around 0.5 [12], the four groups of parameters are chosen including 0.5 to calculate the $\overline{\text{RMS}}$, as shown in Table 4.1. From Table 4.1, we can see that the optimal combination in each group has the lowest value of $\overline{\text{RMS}}$ and the value of n_r is the largest among the three parameters. It implies that the longer wavelengths of photoreceptor cells correspond to the larger n . However, the $\overline{\text{RMS}}$ only reflects the average transformation biases of R, G and B channels, which do not consider the interactions among the three channels. Further analysis of the four optimal combinations is given in the following subsections.

Table 4.1: THE $\overline{\text{RMS}}$ OF FOUR GROUPS OF PARAMETERS, AND EACH GROUP HAS SIX COMBINATIONS. THE OPTIMAL COMBINATION IN EACH GROUP IS HIGHLIGHTED IN BOLD, WHICH HAS THE LOWEST VALUE OF $\overline{\text{RMS}}$.

n_r, n_g, n_b	$\overline{\text{RMS}}$	n_r, n_g, n_b	$\overline{\text{RMS}}$	n_r, n_g, n_b	$\overline{\text{RMS}}$	n_r, n_g, n_b	$\overline{\text{RMS}}$
0.9,0.5,0.1	96.65	0.8,0.5,0.2	99.99	0.7,0.5,0.3	102.13	0.6,0.5,0.4	103.41
0.5, 0.1, 0.9	101.00	0.5, 0.2, 0.8	102.56	0.5, 0.3, 0.7	103.53	0.5, 0.4, 0.6	104.00
0.1, 0.9, 0.5	99.37	0.2, 0.8, 0.5	101.74	0.3, 0.7, 0.5	103.20	0.4, 0.6, 0.5	103.93
0.9, 0.1, 0.5	96.96	0.8, 0.2, 0.5	100.27	0.7, 0.3, 0.5	102.36	0.6, 0.4, 0.5	103.55
0.5, 0.9, 0.1	98.73	0.5, 0.8, 0.2	101.17	0.5, 0.7, 0.3	102.73	0.5, 0.6, 0.4	103.64
0.1, 0.5, 0.9	101.33	0.2, 0.5, 0.8	102.86	0.3, 0.5, 0.7	103.77	0.4, 0.5, 0.6	104.15

4.1.2 Divisive Normalization

Divisive normalization can model the contrast gain control by a nonlinear operator, which is a suitable canonical computational model for processing sensory information underlying adaptation [15], [102], [177]. For normalization of the neural computation, the responses of neurons are divided by a common factor that typically includes the summed activity of a pool of neurons [15]. There are many divisive normaliza-

tion models proposed for achieving gain control in the past twenty years [102], [178], [179]. They mostly aim to establish a connection between the gain control and the statistical properties of natural sensory stimuli [102]. However, how the component like the feedforward component in the divisive normalization model performs better in enhancing low light image data lacks further exploration.

Combining the characteristics of the self-shunting model [12], [180] and the normalization model of the odorant receptor [15], [115], we design a new divisive normalization processor in the proposed dark adaptation. This is because the perceived intensity produced by the power law operation needs to be adaptively mapped into the range of $0 \sim 1$ for further canonical computation. Compared with the traditional normalization model [15], the proposed normalization processor simply utilizes the average intensity of inputs to control the normalization output instead of the power law output of the average intensity. The background light intensity of the input as a controlling factor in the normalization model can better reflect the real illumination conditions.

Here, the designed divisive normalization processor can be mathematically described by,

$$\psi'_R = \frac{I_R^{n_r}}{I_R^{n_r} + I_{m_R}} = \frac{\psi_R}{\psi_R + I_{m_R}} \quad (4.4a)$$

$$\psi'_G = \frac{I_G^{n_g}}{I_G^{n_g} + I_{m_G}} = \frac{\psi_G}{\psi_G + I_{m_G}} \quad (4.4b)$$

$$\psi'_B = \frac{I_B^{n_b}}{I_B^{n_b} + I_{m_B}} = \frac{\psi_B}{\psi_B + I_{m_B}} \quad (4.4c)$$

where ψ'_R , ψ'_G and ψ'_B denote the normalized sensation magnitude output. And, I_{m_R} , I_{m_G} and I_{m_B} are the averaged intensities with respect to the three color channels. They are able to control the response gain as an additive term in the denominator. It is worthy to note that the additive term plays an important role in controlling the normalized outputs, which can be verified by the subfigures in Fig. 4.5.

Fig. 4.5 displays the divisive normalization outputs for the four optimal combinations of parameters that have been mentioned in Section 4.1.1. Given a specific value for I_{m_R} , I_{m_G} and I_{m_B} , the divisive normalization outputs are presented, as shown in

Fig. 4.5(a), (b) and (c). Note that the average intensity of each color channel is set with the same value ($I_{m_R} = I_{m_G} = I_{m_B} = I_m$), and the input I from the three color channels is in the range of $0 \sim 255$. From Fig. 4.5(a) to Fig. 4.5(c), we can see that the normalization output becomes larger and the curve at low values becomes steeper when the average intensity decreases for the four optimal combinations of parameters. This indicates that the divisive normalization operation is effective in raising the intensities of dark pixels. Additionally, we have found that the outputs corresponding to the fourth combination (the last row) have the least variance among the R, G and B channels compared to the other three combinations (rows one to three). Therefore, to keep more consistent of the color components in the final enhanced images, we choose the fourth combination $n_r = 0.6$, $n_g = 0.5$ and $n_b = 0.4$ as useful parameters in our application.

The goal of the divisive normalization processor is to map the result of the compressed data from power law operation to the range of $0 \sim 1$. Although this normalization operation indeed raises intensities for dark pixels, the responses cannot achieve maximized transmission as the produced maximum output is varied when the average intensity changes. To address this problem, the adaptive rescaling operation [181] is required after the divisive normalization operation. More details are described as follows.

4.1.3 Adaptive Rescaling

The rescaling operation can match the dynamic range of responses to that of the inputs when the dynamic range of inputs changes [181]. It ensures the adaptation of visual responses to the mean light level, even though there are fluctuations around the mean. The adaptive rescaling process in the three color channels is mathematically given by,

$$I_c' = \frac{\psi_c' - \min(\psi_c')}{\max(\psi_c') - \min(\psi_c')} \quad (4.5)$$

where I_c' , $c \in (R, G, B)$ denotes the rescaled output for the three color components. It is worthy to note that $\max(x)$ and $\min(x)$ indicate the implementations of the maximum and minimum operation.

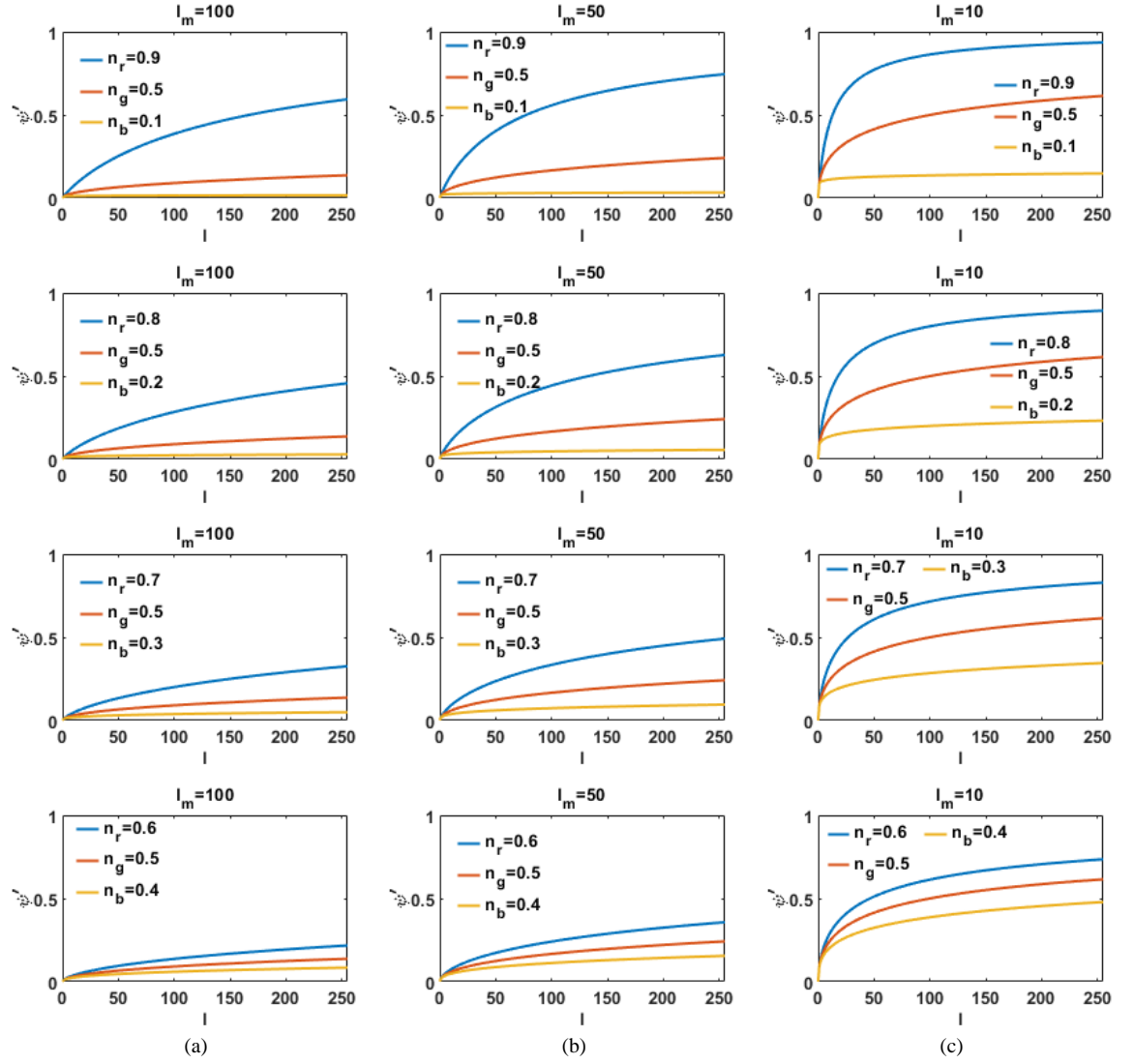


Figure 4.5: The divisive normalization operation for different value of the additive term I_{m_R} , I_{m_G} and I_{m_B} (refer to Eq. 4.4). Here, we set the average intensity in each color channel (R, G, and B) with the same value, $I_{m_R} = I_{m_G} = I_{m_B} = I_m$. (a) $I_m = 100$. (b) $I_m = 50$. (c) $I_m = 10$.

Fig. 4.6 shows the rescaling results under different average intensities, where the contrasts of inputs with low intensities are greatly increased. The rescaling outputs in Fig. 4.6 are shown more adaptive than those without ones (see Fig. 4.5), which indicates that the rescaling operation can maximize the transmitted information. This also explains the reason why it is needed after the divisive normalization processing.

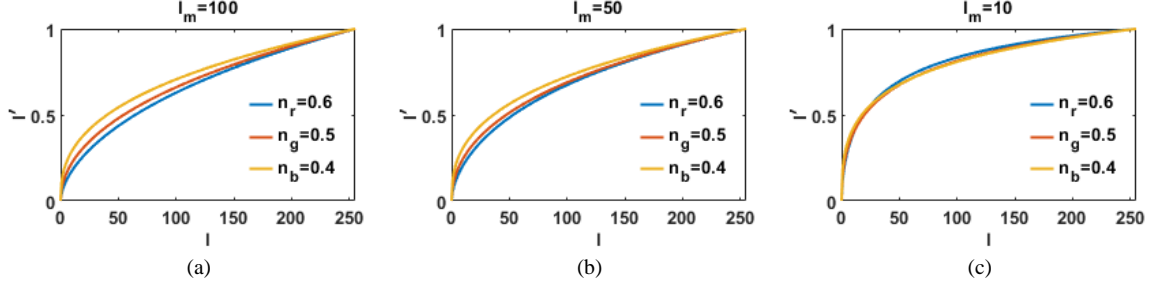


Figure 4.6: The results of adaptive rescaling operation with respect to the adaptation parameters $n_R = 0.6$, $n_G = 0.5$ and $n_B = 0.4$ (refer to Eq. 4.5).

4.2 Experimental Results

1) Data Set: The proposed bio-inspired dark adaptation framework is evaluated on a real low-light image dataset. The dataset contains static images with a resolution of 3000×4000 and images from video clips with a resolution of 960×1280 . All images were captured by a GoPro camera under natural or artificial illumination conditions.

2) Evaluation Metric: For preserving the naturalness of the image during the enhancement processing, no light source should be introduced to the scene, no halo effect should be added and no blocking effect should be amplified [182]. The naturalness of the image is also defined as: the global ambience of the image should not be changed seriously and the direction of the light source should not be altered obviously [135]. Since the naturalness of an enhanced image is related to the relative order of lightness in different local areas [135], we therefore employ lightness order error (LOE) to quantitatively measure the lightness distortion of enhanced results. The lower LOE indicates that the enhancement better preserves the naturalness of illumination [20]. The definition of LOE is mathematically given by [172],

$$LOE = \frac{1}{m} \sum_{x=1}^m RD(x) \quad (4.6)$$

where m is the pixel number, $RD(x)$ is the relative order difference of the lightness between the reference image and the enhanced version of the low-light image for pixel x , which is defined as follows:

$$RD(x) = \frac{1}{m} \sum_{y=1}^m U(Q(x), Q(y)) \oplus (U(Q_r(x)), Q_r(y)) \quad (4.7)$$

where the function $U(p, q)$ returns 1 if $p \geq q$, 0 otherwise, \oplus stands for the exclusive-or operator. $Q(x)$ and $Q_r(x)$ denote the lightness component at location x of the enhanced and reference image respectively, which are the maximum values among R, G and B channels. However, the LOE is highly complex computationally. Hence, a down-sampling is used to reduce the complexity of computing LOE [172]. Specifically, all images are down-sampled to a fixed size 100×100 by collecting 100 rows and columns evenly.

According to the definition of LOE, we can find that its value depends on the lightness component of the reference image (Q_r). In [20], it has pointed out that using the low-light input as the reference is problematic since there exists an extreme case that the LOE is 0 when no enhancement is performed. However, they adopt the HDR [171] result as the reference which is a groundtruth produced by the artificial enhancement algorithm. For the sake of objectiveness, we use the image of the same scene captured under higher illuminations as the reference with respect to the dark and very dark images.

2) Implementation: The proposed bio-inspired dark adaptation framework was implemented in Matlab environment on a standard laptop with a 1.80GHz Intel Core i5 CPU and 16.00GB RAM memory. The adaptation parameters n_R , n_G and n_B are set with 0.6, 0.5 and 0.4, which have been explained in Section 4.1.

4.2.1 Comparison on Real Low-Light Dataset

To evaluate the performance of the proposed bio-inspired dark adaptation framework, we compare it with existing low light image enhancement methods, including Adaptive Contrast-Limited Adaptive Histogram Equalization (CLAHE), Matlab HDR

(High-Dynamic Range) image tone mapping (HDR), Illumination Estimation based method (LIME) [20], Bio-Inspired Multi-Exposure Fusion method (BIMEF) [172] and Gradient-based method (Tanaka's) [21]. Note that the codes of LIME, BIMEF and Tanaka's are downloaded from the authors' websites using default parameters. For CLAHE and HDR, we use the *adapthisteq* and the *tonemap* function integrated in the Matlab toolbox applying default parameters. Particularly, the operation of CLAHE is executed on the L channel by first converting it from the RGB colorspace to the LAB one and then converting the processed LAB back to the RGB colorspace.

The sample images of the tested low-light dataset are shown in Fig. 4.7. There are twelve images recording four different scenes. Each scene includes three images captured under different levels of illumination. Specifically, the images captured under dark and very dark environments are tested images utilized to compare the performance between the proposed method and other enhancement methods, while the images captured under higher illuminations are regarded as the reference images for calculating the LOE.

Fig. 4.8 provides the visual comparison among the competitors on different scenes of the low-light image dataset where the dark inputs are displayed in the second column and the third column in Fig. 4.7. From Fig. 4.8, we can observe that the proposed method has better visual performance when the enhanced image does not implement further denoising processing compared with other competitors, especially in very dark conditions.

Table 4.2 shows the LOE numbers of all the competitors on the low-light image dataset. As can be seen from the numbers in Table 4.2, the proposed method has the lowest average LOE which significantly outperforms others.

Table 4.3 gives the average running time comparison of the six methods on the low-light image dataset where the proposed method costs the least average running time. It demonstrates that the proposed method is much more efficient in computation than other competitors.

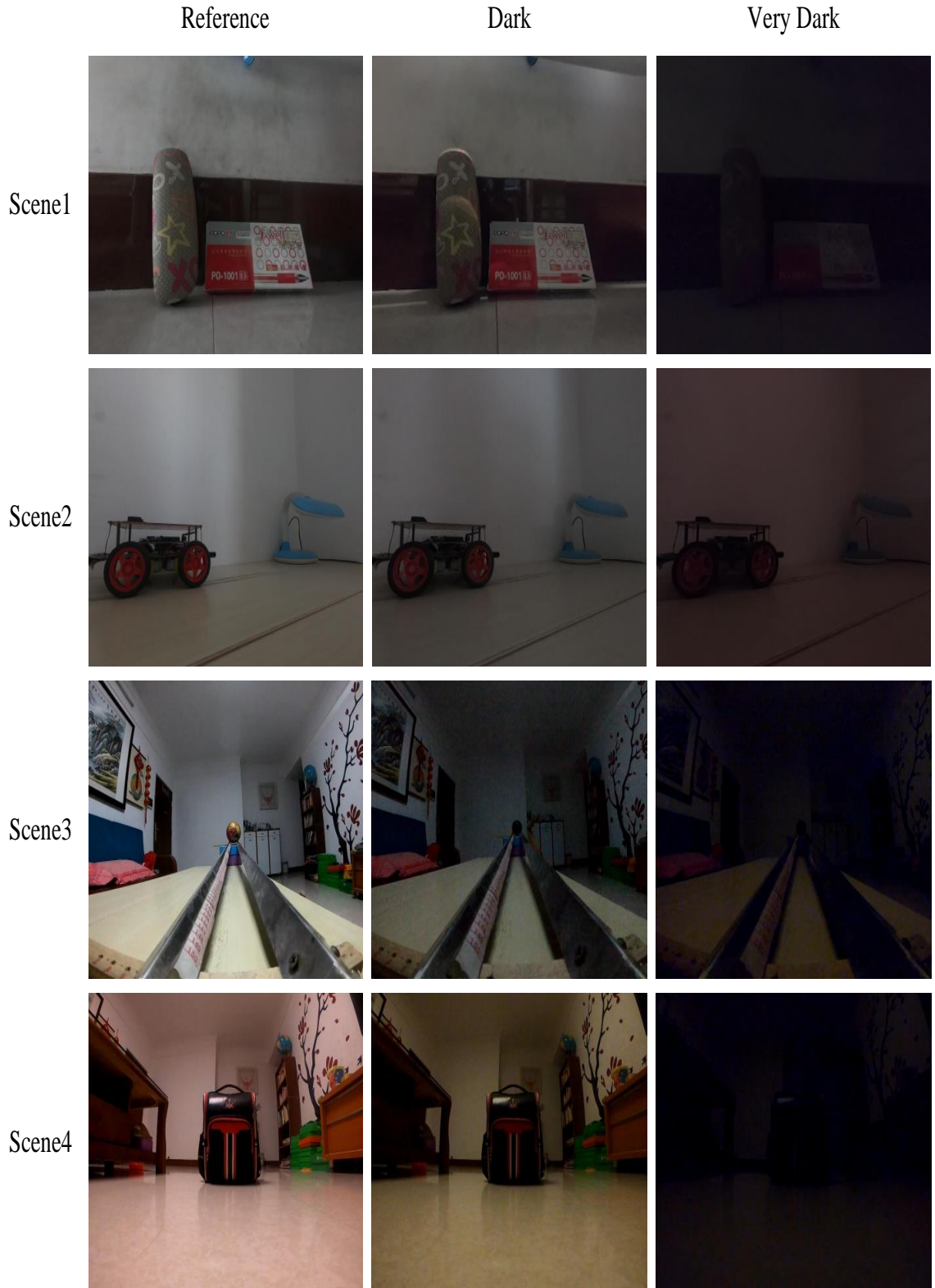


Figure 4.7: Sample images in the low-light image dataset. From the first column to the third column, images are captured under different levels of illumination. The scenes recorded by the images in each row are the same. Specifically, scene1 and scene2 are static images while scene3 and scene4 are images from video clips. Note that images in the first column are captured under higher illumination, which is used in the evaluation experiments as the reference images for the dark and very dark scenes in the second and the third columns.

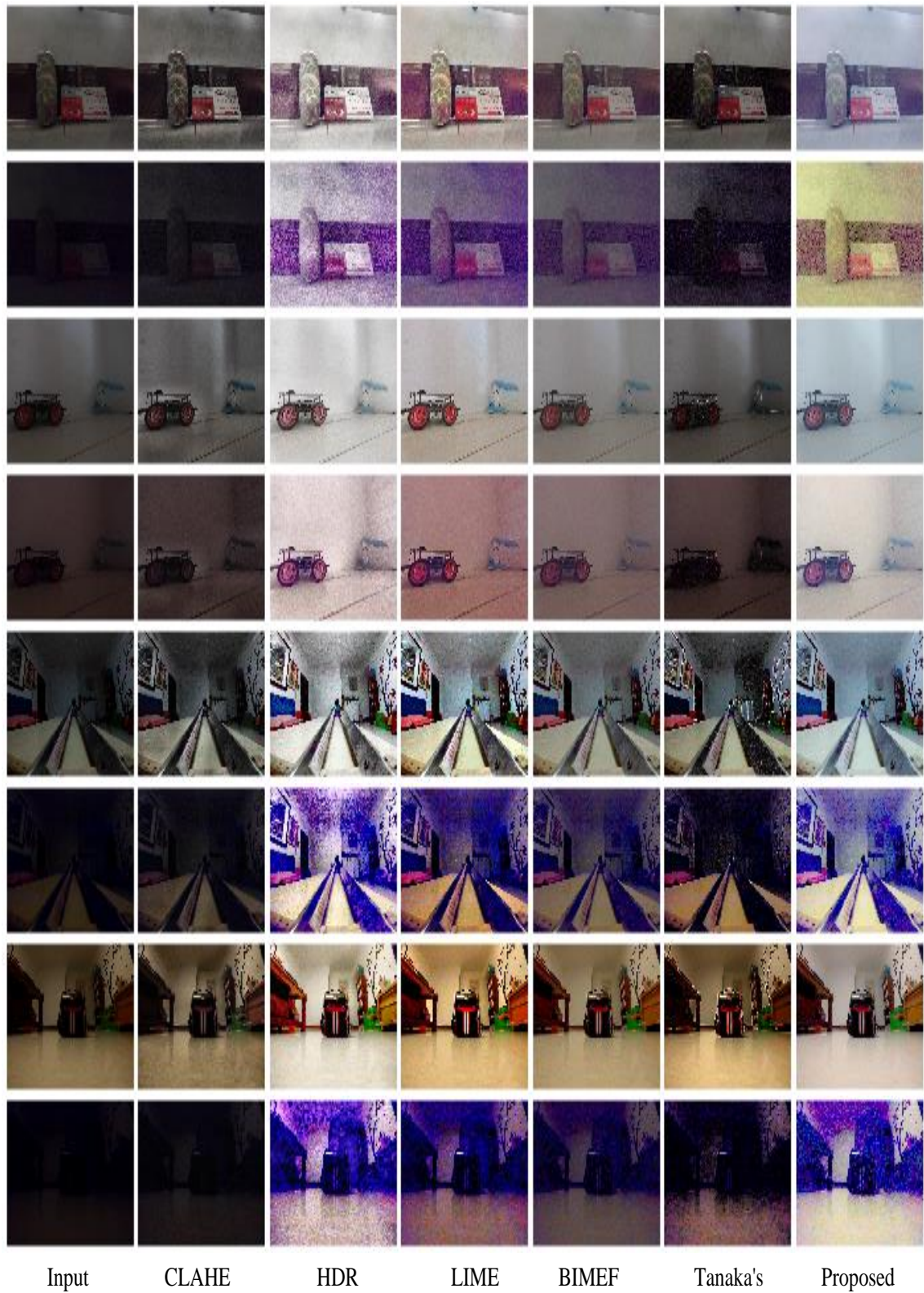


Figure 4.8: Visual comparison among the competitors on the low-light image dataset.

Table 4.2: QUANTITATIVE PERFORMANCE COMPARISON ON THE LOW-LIGHT IMAGE DATASET IN TERMS OF LOE. LOE HAS A FACTOR 10^3 . **THE LOWER THE LOE IS, THE BETTER THE ENHANCEMENT PRESERVES THE NATURALNESS OF ILLUMINATION.**

Method	Scene1 Dark	Scene1 Very Dark	Scene2 Dark	Scene2 Very Dark	Scene3 Dark	Scene3 Very Dark	Scene4 Dark	Scene4 Very Dark	Ave. LOE
CLAHE	2.162	2.808	2.592	1.038	2.216	2.450	2.663	1.426	2.169
HDR	1.858	2.479	2.744	0.988	2.611	1.930	3.420	2.488	2.315
LIME	2.032	2.058	2.741	1.274	2.728	1.913	3.170	2.194	2.264
BIMEF	1.625	1.379	2.474	0.452	2.028	1.171	2.999	1.571	1.712
Tanaka's	1.695	1.493	2.723	0.822	2.675	1.174	3.027	1.856	1.933
Proposed	1.675	1.441	2.392	0.297	1.482	1.001	2.924	1.260	1.559

Table 4.3: AVERAGE RUNNING TIME COMPARISON OF THE SIX ENHANCEMENT METHODS ON THE LOW-LIGHT IMAGE DATASET. THE SIZES OF IMAGES ARE 1280 PIXELS (HORIZONTAL) \times 960 PIXELS (VERTICAL) AND 4000 PIXELS (HORIZONTAL) \times 3000 PIXELS (VERTICAL).

	CLAHE	HDR	LIME	BIMEF	Tanaka's	Proposed
Time (s)	6.160	8.231	4.769	5.294	35.441	1.182

4.3 Discussion

In the above section, we have proposed a bio-inspired dark adaptation framework for low-light image enhancement. The experiments showed that the proposed method performs better visual performance and the results are closer to the references than the others when challenged with low-light images. However, the proposed method only considers raising the intensities but does not deal with the inevitable noise problem. In [11], [116], the bio-inspired night-vision algorithms use spatio-temporal summation neural strategy to reduce noise. Therefore, for low-light image enhancement, it is worth combining other neural strategies together in future work to obtain more pleasant visual performances.

4.4 Chapter Summary

To summarize, we have proposed a low-light image enhancement method which involves dark adaptation processing in three color channels. The key to the dark adap-

tation is to adaptively raise the intensities of dark pixels by integrating a series of canonical neural computations. Moreover, the important parameters for power law adaptation are analyzed in detail for the three color channels. Experimental results on the low-light dataset have demonstrated its effectiveness and efficiency in enhancing intensities and contrasts as well as preserving the naturalness. It therefore can feed the vision-based application like motion detection in low light environments as a positive enhancement method.

In our future work, we will integrate other neural mechanisms with the proposed model for further denoising to obtain better visual performance. Additionally, it can be combined with other techniques for motion detection in low light conditions.

Chapter 5

DLMCE- Dim-light Motion Cues

Enhancement

Moving edges as visual stimuli of motion vision are critical cues for insects' navigation and course control as well as searching for mates, tracking prey, or predator detection [183]. The extraction of motion cues (moving edges) from dynamic visual scenes is one of the earliest and most important processing steps in any biological visual system [37]. It is the specific motion-sensitive neurons that respond to motion cues selectively make animals' behaviors such as locomotion, avoiding imminent collision, tracking moving targets, etc possible [40], [51], [153], [156], [184]. For example, in locusts, looming cues are signaled by the lobula giant movement detector neurons (LGMDs) [29], [36] that have been modeled for collision detection [7], [156], [185]. Therefore, the detection of visual motion cues is essential for an artificial vision system to perceive visual motion in a timely and robust manner.

In the real world, on the other hand, an artificial vision system has to be able to cope with various lighting conditions equally well - not only in the daytime with sufficient light but also at dawn or nighttime with dim light conditions. It is a challenging task for many artificial vision systems to detect precise motion cues in very low light conditions [168], [186]. Motion cues captured by image sequences are regarded as spatio-temporal luminance changes [142]. Also, they can be visualized as moving edges of an object with a certain thickness depending on visual motion velocities. At high levels of ambient light, the motion cues can be successfully detected due to high

signal-to-noise ratios (SNRs) [32], [38], [39], [41]. When light levels are low, camera sensors cannot reliably measure the pixel intensity of the visual world due to the scarcity of photons. This is why the low-light images usually accompany extremely low SNRs as well as the very low image contrast. The motion cues are correspondingly very weak. Hence, enhancing motion cues is vital for the successful detection of moving objects in dark environments.

For the enhancement of dim-light motion cues, an intuitive solution is to enhance the low light image directly. Traditional enhancement methods for low-light images include intensity transformation and denoising processing as raising intensity also inevitably amplifies noises [19], [120], [130], [187]–[189]. However, these methods only improve the visual quality of the dark images, and the improvement for motion cues is very limited. Indeed, a “bad” enhancement will lead to even worse results of motion cues as the motion information may be broken after enhancement. Moreover, motion cues detection, especially collision cues detection for autonomous systems like robots or vehicles, requires a real-time response to a looming object - the traditional image enhancement and denoising algorithms for low light images are usually too computationally expensive [10], [188], [189] to run in real-time.

In the real world at night times, nocturnal insects show remarkable visual abilities in detecting faint movements with their compact visual brains [27], [28]. Their exquisite visual systems provide abundant sources of inspiration for the development of the functionality of motion cues enhancement. This biological solution for enhancing motion cues makes it possible to develop an efficient and fast motion detection system. Existing biological methods for night vision focus on the enhancement of dark images based on the widely investigated neural summation strategy [10], [11], [42], [43]. However, these approaches are devoted to constructing adaptive spatio-temporal smooth kernels to reduce the noise and preserve the edge, while the structure tensor kernel is computationally complex and time-consuming consequently. How exactly neural strategies in biological visual systems perform night vision so effectively with a limited number of neurons remains unclear. A more efficient way of neural information processing needs to be explored for enhancing the motion cues in dim light

environments.

Inspired by the visual systems of nocturnal animals, we propose a new bioinspired neural network model that enhances only the faint motion cues in dim light conditions by adopting canonical neural computation [15] and neural summation [17] strategies. The proposed dim-light motion cues enhancement model (DLMCE) firstly utilizes the dark adaptation mechanism to increase the image intensity and contrast. The prominent motion cues can then be obtained by the spatio-temporal constraint mechanism. Finally, these motion cues are further enhanced by the neural summation of lateral excitation and self-inhibition mechanisms. The main contribution of this chapter is to combine canonical neural computations and neural summation mechanisms to extract motion cues from dim-light environments. Compared to existing methods, the proposed DLMCE model is effective and efficient in enhancing faint motion cues as it focuses on recognizing and enhancing real moving edges without complex denoising processing.

The remainder of this chapter is organized as follows. Section 5.1 introduces the formulation of the proposed DLMCE model. Section 5.2 reports comprehensive experimental results as well as performance comparisons with comparative methods on both synthetic and real datasets, including the combination of the motion cues detection methods and conventional low-light image enhancement methods. Finally, we conclude this chapter in Section 5.3.

5.1 Formulation of the DLMCE Model

The proposed DLMCE model is shown in Fig. 5.1, which is composed of six processing layers from the retina to the lamina. In the retina, the L layer firstly implements the dark adaptation operation. Then, the P layer calculates the luminance changes as potential motion cues, whilst using the spatio-temporal constraint strategy to extract prominent motion cues. Finally, they are passed into the lamina where the E, I, S and M layers utilize neural summation strategies to further enhance the extracted motion cues from the P layer. Fig. 5.2 shows the architecture of the DLMCE, involving signal processing in the retina and lamina. More details are as follows.

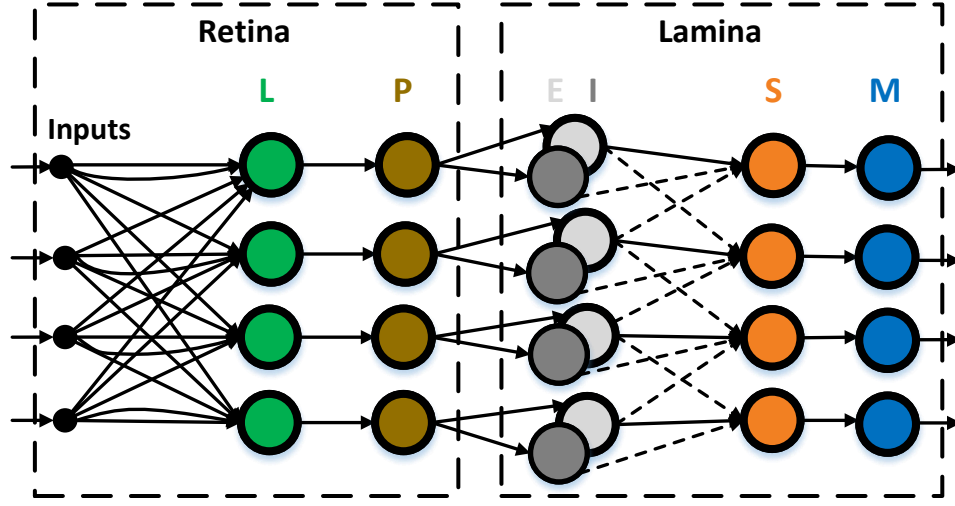


Figure 5.1: The proposed DLMCE model is composed of six layers of cells (L, P, E, I, S, M) from the retina to the lamina. The input to the L cell corresponds to the pixel's luminance. The P cell perceives the changes in luminance and passes them to the E/I cell. The S cell accumulates signals from the E and I cells, including lateral excitation and self-inhibition. Note that the signal without time delay is indicated with the solid line, while the delayed one is indicated with the dashed line. The output signal from the M cell is the enhanced motion cue.

5.1.1 Retina

1) L layer (Dark adaptation): In low light conditions, visual inputs perform very low intensities and low contrasts due to the scarcity of photons. The dark adaptation can adaptively raise the intensity and increase the contrast. It consists of a series of canonical neural computations, involving the power law adaptation [108], the divisive normalization [15], [115] and the adaptive rescaling operation [181], [190]. The dark adaptation processing can be mathematically defined as

$$I_m(t) = \frac{1}{N} \sum_{x=1}^r \sum_{y=1}^c I(x, y, t) \quad (5.1)$$

$$L(x, y, t) = I^n(x, y, t) \quad (5.2)$$

$$L'(x, y, t) = \frac{L(x, y, t)}{L(x, y, t) + I_m(t)} \quad (5.3)$$

$$L_o(x, y, t) = \frac{L'(x, y, t) - L'_{min}}{L'_{max} - L'_{min}} \quad (5.4)$$

where $I_m(t)$ indicates the average intensity of the input image. N is the total number of image pixels that equals rows (r) multiplied by columns (c). I^n represents the

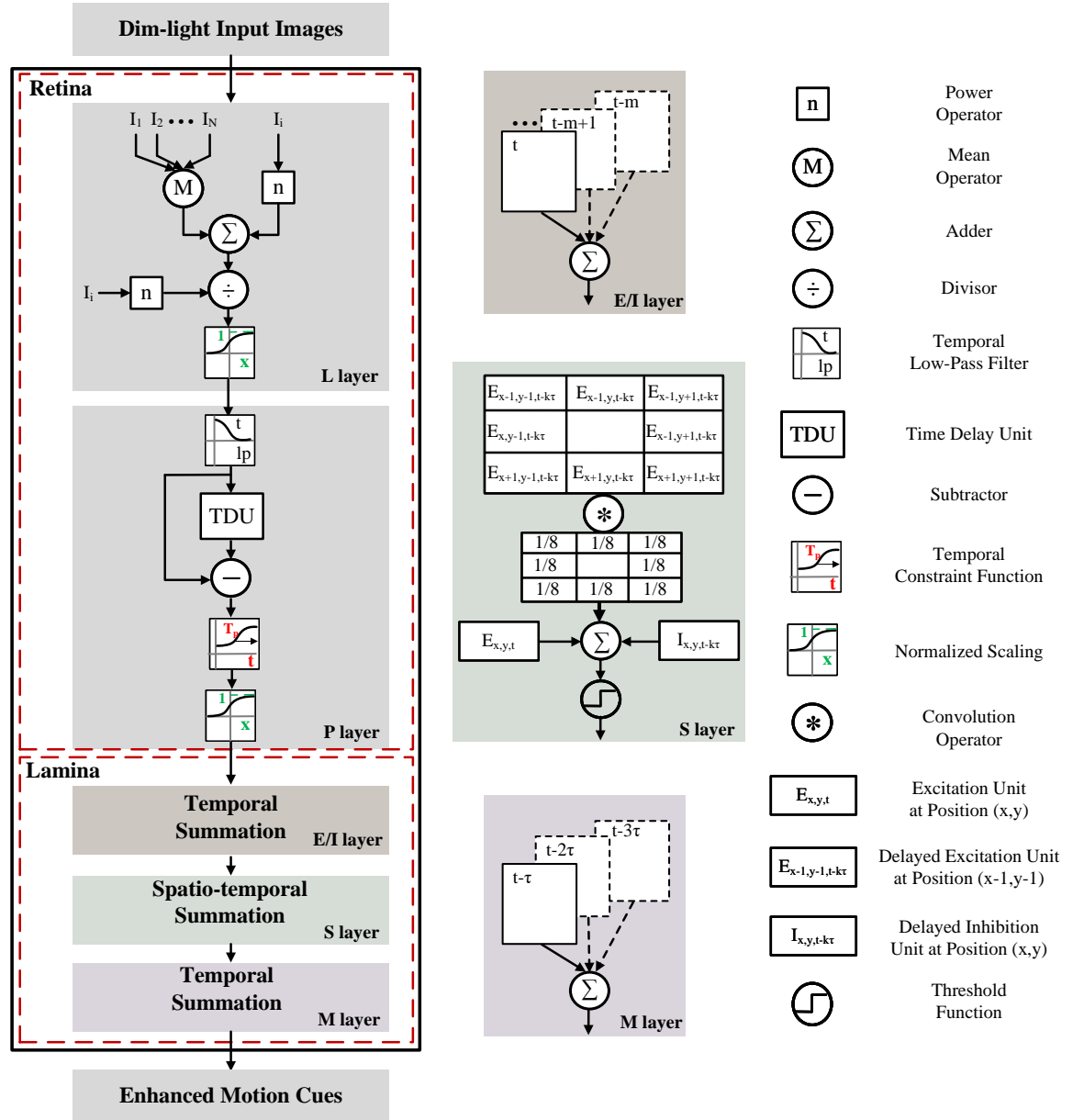


Figure 5.2: Schematic illustration of signal processing in the proposed DLMCE model. The pipeline of signal processing is shown in the left column with the lamina layers explained in further detail in the middle column. In the retina: firstly, the dark adaptation processing in the L layer is responsible for enhancing the dim-light image's intensity and contrast; then, the P layer calculates the luminance changes and implements spatio-temporal constraints for obtaining strong cues. In the lamina: the signal processing is the neural summation, including temporal summation in the E/I layer, spatio-temporal summation in the S layer and temporal summation in the M layer; specifically, the spatio-temporal summation involves the summation of lateral excitations and self-inhibition. Faint motion cues from dim-light input images are extracted and highlighted by the canonical neural computations in the retina and further enhanced by the neural summation processing in the lamina.

power operation for the input image I . Note that the value of I is within the range of $0 \sim 255$. The exponent n is related to the sensory system which is set less than one for fitting intensity/response functions of invertebrate photoreceptors [12]. L'_{max} and L'_{min} denote the maximum and the minimum value of L' respectively. L_o represents the final enhanced luminance results from the dark adaption processing.

2) P layer (Spatio-temporal constraint): The photoreceptor cell in the P layer receives the changes in luminance that last in a specific time interval. That is,

$$P(x, y, t) = abs(\tilde{L}_o(x, y, t) - \tilde{L}_o(x, y, t - \tau)) \quad (5.5)$$

$$\tilde{L}_o(x, y, t) = \frac{1}{\int_{t-T}^t \omega(i) di} \cdot \int_{t-T}^t \omega(T-i) \cdot L_o(x, y, i) di \quad (5.6)$$

$$\omega(t) = e^{-t/\sigma_t} \quad (5.7)$$

$$T = n_f * \tau \quad (5.8)$$

where $\tilde{L}_o(x, y, t)$ denotes the filtered luminance signal processed by the low-pass filter; $P(x, y, t)$ represents the luminance changes (increments or decrements); τ represents one frame time interval; σ_t denotes the decay time constant; the time interval T equals the number of frames n_f multiplying frame intervals τ .

Although changes in luminance are regarded as potential motion cues, there are a lot of noises due to extreme low SNR. To retain strong cues and eliminate small and isolated noises, spatio-temporal constraints are therefore implemented. Firstly, the luminance changes are constrained in the temporal domain, which is set to zero by thresholding if they are lower than a specific proportion of the average luminance level at the current time. This temporal constraint process can be mathematically described as

$$\tilde{P}(x, y, t) = \begin{cases} P(x, y, t) & \text{if } P(x, y, t) \geq T_p \\ 0 & \text{otherwise} \end{cases} \quad (5.9)$$

where T_p is an adaptive threshold which is relevant to the mean value of the filtered

luminance signal \tilde{L}_o . That is,

$$T_p(t) = \alpha \cdot \frac{\sum_{x=1}^r \sum_{y=1}^c \tilde{L}_o(x, y, t)}{N} \quad (5.10)$$

where the coefficient α is an adjustable value, which implies that the adaptive threshold T_p is further tuned by the α to obtain strong motion cues. Secondly, the output of the P cell at time t is also constrained by the normalized scaling processing in the spatial domain. That is,

$$\tilde{P}_o(x, y, t) = \frac{\tilde{P}(x, y, t)}{\tilde{P}_{max}} \quad (5.11)$$

where \tilde{P}_{max} denotes the maximum value of \tilde{P} in the spatial domain at time t . Note that the minimum value of \tilde{P} equals zero indicating no changes in luminance.

Fig. 5.3 illustrates processes of the aforementioned spatio-temporal constraints. The temporal constraint process is displayed in Fig. 5.3(a), which can inhibit small fluctuations of input luminance while retaining the prominent luminance changes. Meanwhile, the process of spatial constraint is exhibited in Fig. 5.3(b). From Fig. 5.3(b), the overall luminance varies at different time but keeps the same output range ($0 \sim 1$) by the spatial normalization processing. We found that the strong luminance changes in the visual scene at any time can be retained in spite of the change in the overall luminance. As a result, the spatio-temporal constraint mechanism is able to guarantee stable extraction of motion cues even if there are small luminance fluctuations as well as the overall luminance changes.

5.1.2 Lamina

1) E/I layer (Temporal summation): In the E/I layer, an excitatory cell (E unit) and an inhibitory one (I unit) are at the same retinotopic position. They are responsible for collecting the averaged motion cues in a specific time interval. Their outputs are obtained from the temporal summation, which is given by,

$$E(x, y, t) = \frac{1}{m} \cdot \int_{t-m}^t \tilde{P}_o(x, y, i) di \quad (5.12)$$

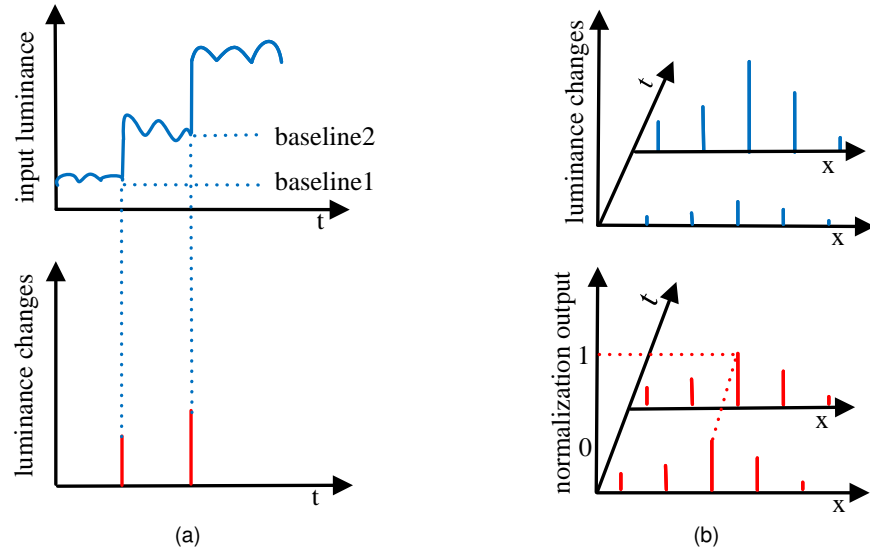


Figure 5.3: Illustration of the spatio-temporal constraint on luminance changes. (a) It displays the temporal constraint processing which exerts an adaptive threshold on changes in input luminance. The baseline1 and baseline2 represent the adapted average intensity value at different times. The extracted luminance changes indicate that small fluctuations of input luminance are well inhibited. (b) It exhibits the spatial constraint processing where the normalization is implemented for the inputs. Outputs in the x axis at different instants in time t are constrained in a fixed range ($0 \sim 1$), which ensures stable outputs in spite of the overall luminance changes caused by different light levels. Note that the x coordinate represents the spatial position of image pixels.

Note that the output of the I cell is identical to the E cell.

2) S layer (Spatio-temporal summation): The extracted motion cues from the E/I layer are further enhanced by the S layer's spatio-temporal summation mechanism which is illustrated in Fig. 5.4. As can be seen from Fig. 5.4, the excitatory flow

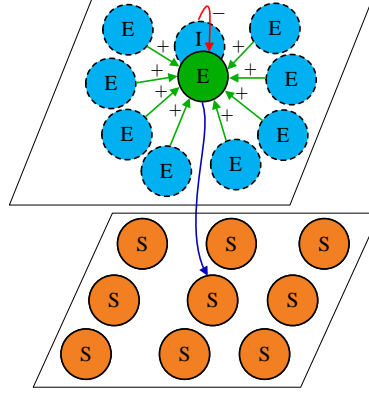


Figure 5.4: Illustration of S layer's spatio-temporal summation mechanism. The excitatory (E) unit at the center position adds excitations from neighboring E units and the inhibition from the inhibitory (I) unit. Its output is then passed to the summation (S) unit in the same retinotopic position. The cell that transfers the signal with time delay is marked with dashed lines. The signals of delayed lateral excitations are indicated by the green arrows whereas the delayed self-inhibition signal is indicated by the red arrow. For clear illustration, only one S cell's summation processing is shown here.

from the E cell and its neighboring E cells, as well as the inhibition from the I cell are summed by the S cell. It is worthy to note that the lateral excitations and the inhibition are passed to the S layer by neighboring E cells and I cell after a delay of a specific time. Since the I unit has the same retinotopic position as the E unit, the inhibition signal can be regarded as self-inhibition. The spatio-temporal summation of the S cell is mathematically described as follows.

$$E_N(x, y, t - k\tau) = \sum_{i=-1}^1 \sum_{j=-1}^1 E(x + i, y + j, t - k\tau) w_e(i, j) \quad (5.13)$$

$$w_e = \frac{1}{8} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (5.14)$$

$$S_{k\tau}(x, y, t) = E(x, y, t) + E_N(x, y, t - k\tau) - I(x, y, t - k\tau) \quad (5.15)$$

where $E_N(x, y, t - k\tau)$ represents neighboring excitations of pixel position (x, y) with delays of k frame intervals. w_e is the connecting weight matrix with respect to the neighboring excitation to the current position. $I(x, y, t - k\tau)$ denotes the self-inhibition with the same time delays. $S_{k\tau}(x, y, t)$ indicates the output of the S cell, which is the enhanced motion cues. If the value of the S cell exceeds the threshold T_s , then the S cell is fired representing there is a useful motion cue. Otherwise, the response of the S cell is set to zero. That is,

$$S_{k\tau}(x, y, t) = \begin{cases} 1 & \text{if } S_{k\tau}(x, y, t) \geq T_s, k \in Z^+ \\ 0 & \text{otherwise} \end{cases} \quad (5.16)$$

where τ is the frame interval and the value of k equals the number of the frame.

The spatio-temporal summation processing is able to detect motion cues with various velocities due to the accumulated delayed neighboring excitations. Moreover, it can suppress flickering since the delayed self-inhibition reduces the changes in luminance at the same position which are not real motion cues. The real motion occurs not only with respect to luminance changes but also shifts of positions. Also, the summation in temporal (temporal smoothing) was verified to have the ability in suppressing the flickering noise [119].

3) M layer (Temporal summation): In the M layer, the M cell collects the motion signals passed from the S cell with different time delays at instant t . The output of the M cell is given by:

$$M(x, y, t) = \frac{1}{3} \cdot \sum_{k=1}^3 S_{k\tau}(x, y, t) \quad (5.17)$$

where the value of $M(x, y, t)$ indicates the averaged motion cues in the time duration $(0, 3\tau)$.

5.1.3 Parameters of the System

The parameters of the proposed DLMCE model are listed in Table 5.1 based on current trials in this chapter. The parameters of σ_t , n_f , and T_s are determined empirically.

ically. The parameter τ is determined by the frame rate. It is worthy to note that the parameter n is set less than one since it fits the dark-adapted intensity/response functions of invertebrate photoreceptors best [12]. To further explain the determination of the $n = 0.4$, we conduct relevant experiments and analysis which is reported in Section 5.2.2. In the following experiments, they will keep unchanged unless stated. Note that the given parameter T_s is a threshold that can make the model show preference for strong motion cues.

Table 5.1: PARAMETERS OF THE PROPOSED DLMCE MODEL

Eq.	Parameters
(5.1)	$n = 0.4$
(5.7)	$\sigma_t = 4$
(5.8)	$n_f = 3, \tau = 8.33 \text{ ms}$
(5.10)	$\alpha = 0.2$
(5.16)	$T_s = 0.35$

5.2 Experimental Results

5.2.1 Experimental Setup

1) Data Sets: We have evaluated our DLMCE model on synthetic and real datasets. The synthetic datasets are produced by the Vision Egg software [164], which include static images (see Section 5.2.2) and video sequences (see Section 5.2.3 and Section 5.2.4). Moreover, all synthetic datasets are produced with a resolution of 96×128 . The real dim-light datasets are captured by a GoPro camera, including static images with a resolution of 3000×4000 and video sequences with a resolution of 960×1280 . These real video images are resized to 96×128 before feeding to the model by using the ‘imresize’ function in Matlab image processing toolbox. The video images were all taken at 120 frames per second (fps). Both synthetic and real sequences involve translating/looming motion under dim or very dim scenarios.

2) Evaluation Protocol: To evaluate motion cues enhancement in dim-light environments, we are concerned with the number of detected real motion cues (N-cues), precision and recall. This is because more motion cues are helpful in the successful

detection of a moving object in a dim light environment and motion detection is the ultimate goal of motion cues enhancement. Moreover, precision is able to measure the fraction of true positives in the motion cues produced by a detector while recall can measure the fraction of ground-truth motion cues detected [191]. Given a binary ground truth and a detection result of motion cues, we can get the number of real motion cues as well as the precision and recall. Since motion cues are moving pixels within a motion trajectory in a 3D spatiotemporal space, the number of the ground truth of cues is related to the length of time (consecutive frames). As a result, the different number of consecutive frames will produce different results. Here, we set the length of time five frames for balancing the result of N-cues. Additionally, the global F-measure is defined as the harmonic mean of precision and recall, which is commonly used to quantitatively evaluate algorithms in previous studies [18], [192].

In this chapter, we employ N-cues and F-measure to evaluate our model since they can reflect the quantity and quality of the detected motion cues. Let G_t be the reference cues map corresponding to the ground truth and D_c the detected cues map at a specific time instant. Afterwards, G_t and D_c are combined, denoting $|\cdot|$ as the cardinality of a set. Then, the N-cues and the F-measure are given by:

$$N-cues = |D_c \cap G_t| \quad (5.18)$$

$$F-measure = 2 \times \frac{precision \cdot recall}{precision + recall} \quad (5.19)$$

where $precision = \frac{|D_c \cap G_t|}{|D_c|}$ and $recall = \frac{|D_c \cap G_t|}{|G_t|}$. It is worthy to note that if the number of N-cues equals zero, then the $F-measure$ is set zero.

3) Implementation: The DLMCE model consists of neural computations in the retina and lamina. The aim of the model is to enhance dim-light motion cues in low light conditions. Here, we have tested the principal characteristics of the DLMCE model and have compared the performance of dim-light motion cues enhancement with existing methods on synthetic and real datasets. The implementation of our experiments is as follows: we first verified the effectiveness of the proposed model's dark adaptation mechanism in enhancing intensity and contrast, including showing en-

hanced intensity and contrast outputs as well as visual performance comparison with several conventional low-light image enhancement methods; we then verified the effectiveness of the motion cues enhancement by testing a number of synthetic dim-light image sequences with motions; we finally compared the DLMCE model with existing methods in the detection of motion cues on synthetic and real datasets, including the motion cues detection with/without low-light image enhancement methods.

5.2.2 Effectiveness of the Dark Adaptation

The L layer's dark adaptation processing in the DLMCE model can increase image intensities and contrasts. It consists of a set of canonical neural computations. To validate the effectiveness of the dark adaptation, we conduct three series of experiments from the aspect of the relation between input/output, intensity/contrast enhancement and visual performance. We also compare with the conventional enhancement methods CLAHE [19], LIME [20] and Tanaka's [21] in the second and the third series of experiments. Note that the codes of LIME and Tanaka's are downloaded from the authors' websites using default parameters. For CLAHE, we use the *adapthisteq* function integrated in the Matlab toolbox. Particularly, the parameter applied in CLAHE is greater than the default parameter for achieving better contrast enhancement.

The first series of experiments presented outputs of each stage during dark adaptation processing with respect to the adaptation parameter, whilst inputs are in the range of $0 \sim 255$. As can be seen from Fig. 5.5, the output intensity is greatly increased when the input intensity is at a very low level. These results conform to the biological evidence [12] that the value of power n in the range of $0.4 \sim 0.6$ is the best dark-adapted fitted. Since the curves of L_o are more consistent when $n = 0.4$ than $n = 0.5$ or $n = 0.6$, we set the parameter n equals 0.4 in the following experiments.

In the second series of experiments, we tested the effectiveness in terms of enhancing intensity/contrast on a synthetic dataset. The dataset contains 25 sinusoidal grating images which have different average intensities I_m (see Equ. (5.1)) and root mean square (RMS) contrasts. Specifically, the RMS contrast (C_{rms}) is defined as the

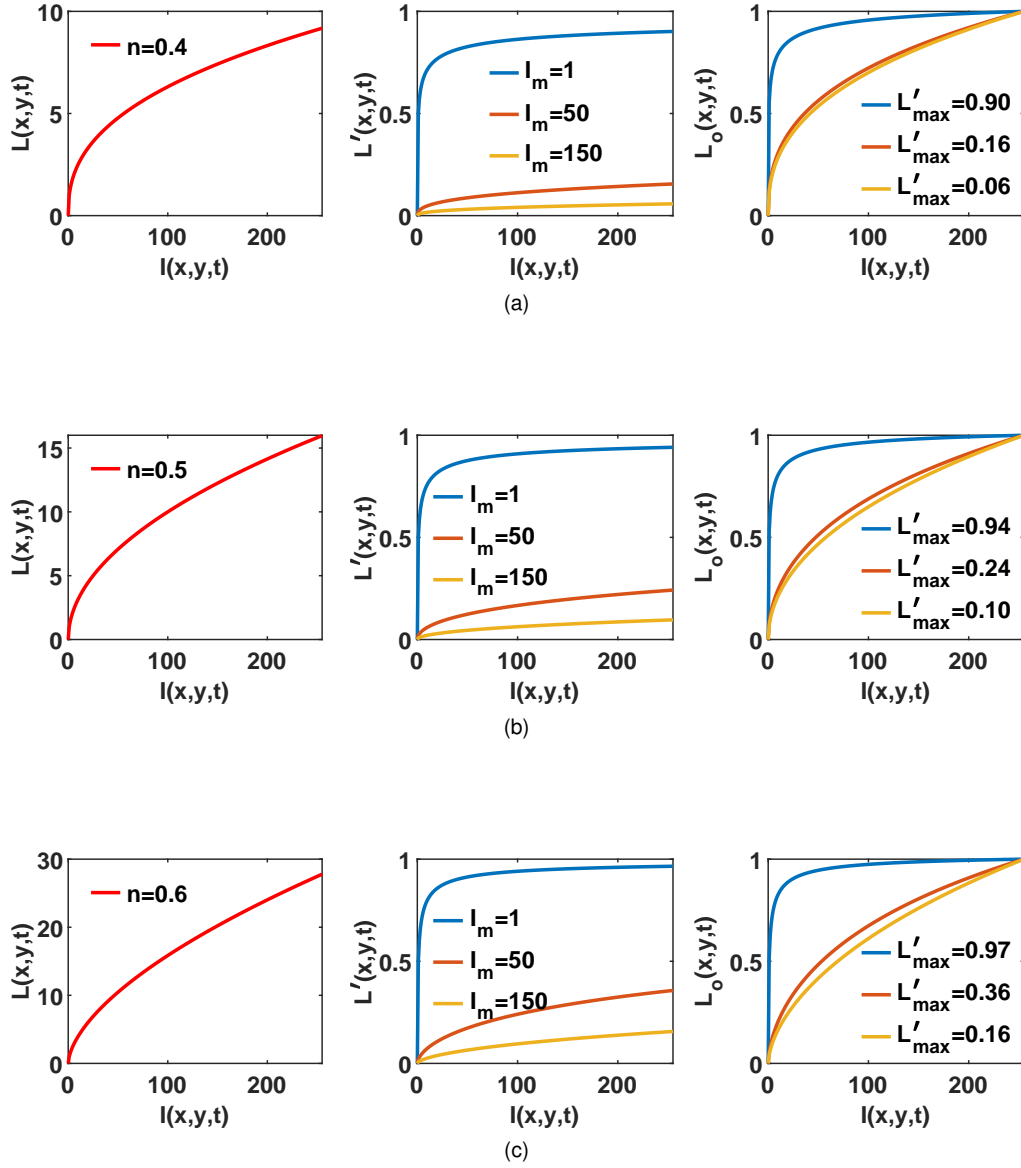


Figure 5.5: (a)-(c): Various outputs during dark adaptation's processing with respect to the adaptation parameter. In each subplot, the horizontal axis denotes input intensity $I(x,y,t)$ while the vertical axis from left to right represent the output of Equ. (5.2), Equ. (5.3) and Equ. (5.4) respectively. In the left column, (a) $n = 0.4$, (b) $n = 0.5$, and (c) $n = 0.6$. Note that the value of I_m is manually set to 1, 50 and 150 individually representing different levels of background luminance. $L_o(x,y,t)$ is the final output for dark adaptation processing.

standard deviation of pixel intensities normalized by the mean [102], i.e.,

$$C_{rms} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (I_i - I_m)^2}}{I_m} \quad (5.20)$$

where N denotes the number of pixels, I_i indicates the pixel's intensity value. These synthetic sinusoidal grating images as inputs are shown in Fig. 5.6(a). Fig. 5.6(b)-(e) are enhanced outputs by various enhancement methods. As can be seen from Fig. 5.6 and Fig. 5.7, the proposed DLMCE's dark adaptation processing is able to enhance the low-light sinusoidal grating images significantly. Particularly, the dark adaptation is superior to other enhancement methods in increasing C_{rms} contrasts for those sinusoidal grating images with very low average intensities (see Fig. 5.7). Note that outputs in Fig. 5.7(d) where input images with average intensity 127 show a slight decrease in C_{rms} contrast. It indicates that the proposed dark-adaptation method is more effective when the average luminance is at a lower level.

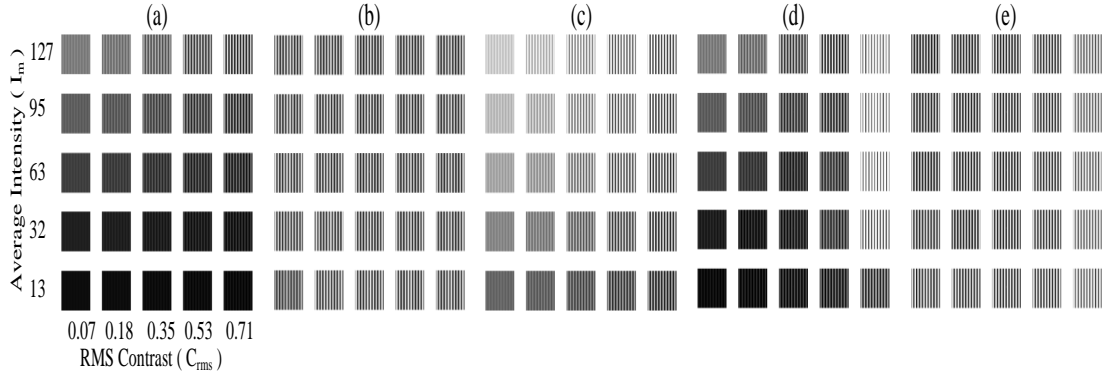


Figure 5.6: Sinusoidal grating images with different average intensity and/or RMS contrast as inputs, as well as outputs by using various image enhancement methods. (a) Input images. (b) CLAHE [19]. (c) LIME [20]. (d) Tanaka's [21]. (e) Proposed DLMCE's dark adaptation.

In the third series of experiments, a real dim-light dataset that contains static images was utilized to test the visual performance. Specifically, the tested images were captured under different low lighting conditions with respect to the same scene. The ambient level of luminosity was measured by a digital light meter represented by lux units. As can be seen from Fig. 5.8(a), the four dim-light input images were captured under 0.4, 0.8, 1.4 and 4.0 lux respectively. Note that the shown input images are grayscale image data converted by the 'rgb2gray' function in the Matlab image processing tool-

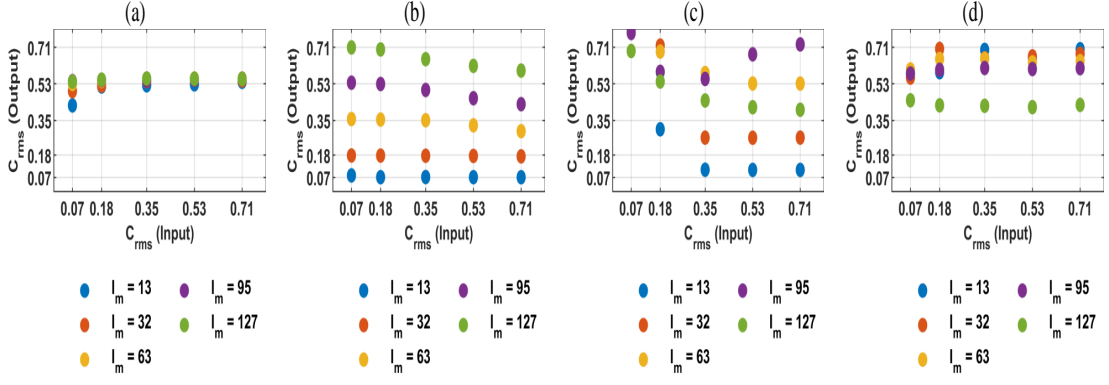


Figure 5.7: RMS contrasts of the input images (see Fig. 5.6(a)) are plotted against the RMS contrasts of the outputs with (a) CLAHE [19], (b) LIME [20], (c) Tanaka's [21], and (d) Proposed DLMCE's dark adaptation.

box. Fig. 5.8(b)-(e) show the visual comparison for the four dim-light inputs, from which, we can find that the results obtained by the proposed DLMCE's dark-adaptation are more visually pleasant. Compared with other enhancement methods, the DLMCE's dark adaptation processing did not introduce too many noises, which is beneficial to the following motion cues enhancement.

5.2.3 Effectiveness of the Motion Cues Enhancement

For insects' visual systems, the ability to perceive motion in dim-light environments is essential for their survival. The extraction of visual motion cues has also been the focus of intense research in motion detection [142]. Obviously, enhancing dim-light motion cues is very important for detecting moving objects in low light conditions. Our proposed DLMCE model enhances motion cues utilizing the spatio-temporal constraint and neural summation mechanisms on the basis of the dark adaptation mechanism. The enhancement procedure of motion cues is illustrated in Fig. 5.9. As can be seen from Fig. 5.9, the extracted motion cues are still weak even though the dark adaptation processing has greatly increased the low-light image's intensities and contrasts while these cues can be further enhanced by the subsequent processing of spatio-temporal constraint and neural summation.

To verify the effectiveness in enhancing motion cues, we tested the proposed model on nine synthetic dim-light image sequences. The sample images from the nine input image sequences are shown in Fig. 5.10(a), which contain a looming black disk. In Fig.

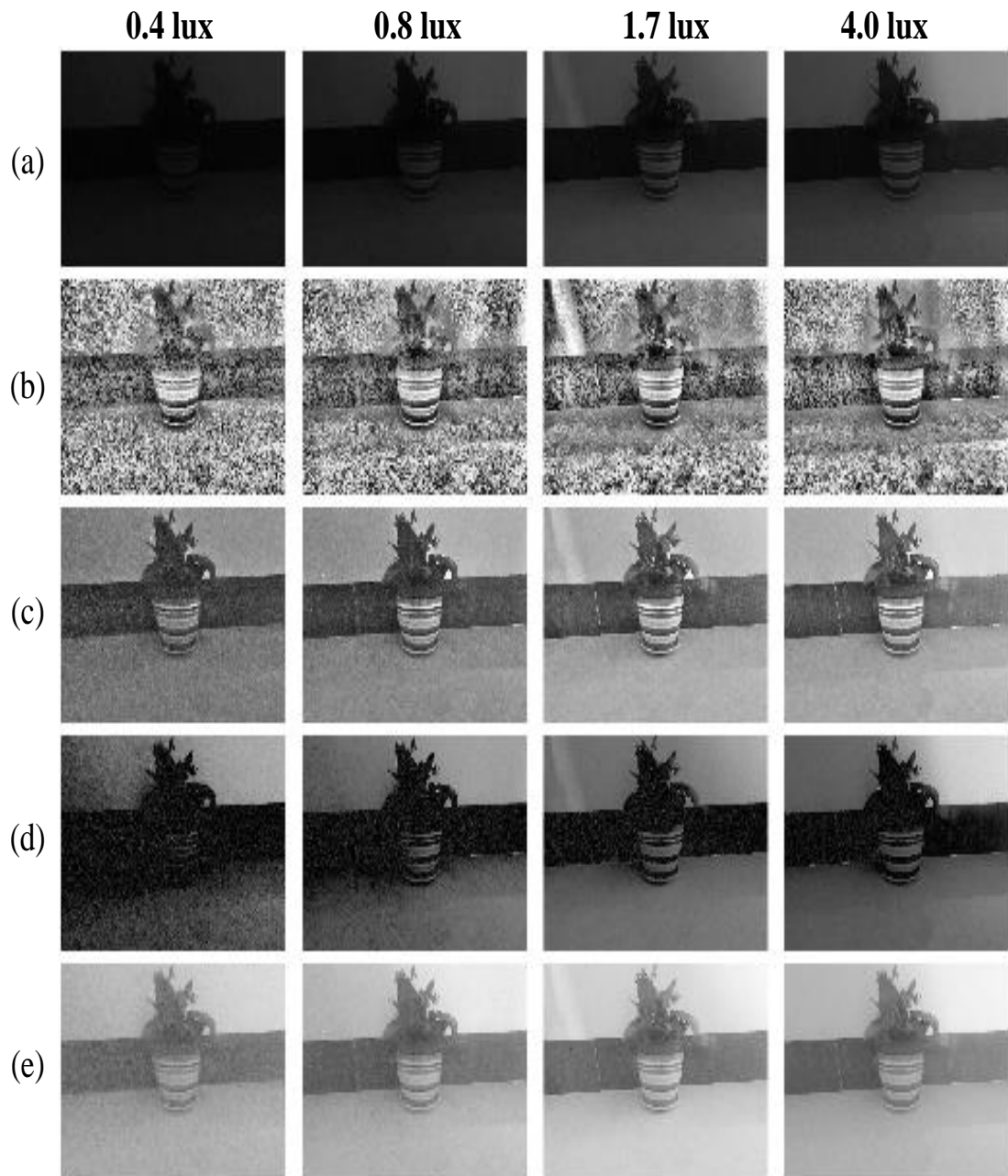


Figure 5.8: Four images of the same scene with different levels of luminosity and corresponding output images by using various image enhancement methods. (a) Dim-light images as inputs. (b) CLAHE [19]. (c) LIME [20]. (d) Tanaka's [21]. (e) Proposed DLMCE's dark adaptation.

5.10(a), the average luminance of backgrounds range from 13 \sim 63 whilst the RMS contrast over a range of 0.07 \sim 0.35. Fig. 5.10(b)-(d) show the enhanced images, motion cues and the detected moving edges respectively. The results have demonstrated that the faint motion cues can be effectively enhanced by the proposed DLMCE model. Note that the effectiveness of enhancement in diminishing static flickering noise by the spatio-temporal summation (see Section 5.1.2) did not be visualized during the motion cues enhancement. This is because the effect of reducing the flickering noise is very hard to display [119].

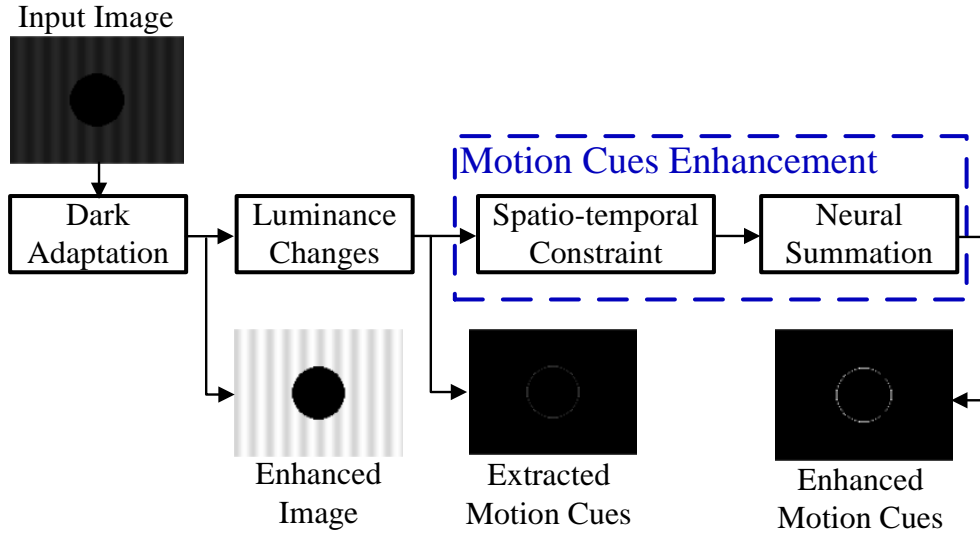


Figure 5.9: Schematic illustration of the enhancement procedure of motion cues in the DLMCE model: a dim light image, as input, is fed into the L layer of the DLMCE model implementing dark adaptation processing; the changes in luminance are computed for extracting motion cues; these extracted motion cues are further enhanced by the subsequent spatio-temporal constraint and neural summation processing.

5.2.4 Comparison on Synthetic and Real Datasets

To further verify the effectiveness, we compared the DLMCE model with comparative methods on synthetic and real datasets. Since there is no published motion cues enhancement model available that can be used to compare at the present, the experimental comparisons are implemented by comparing with motion cues detection methods as well as their combination with low-light image enhancement methods. Specifically, the motion cues detection methods are temporal differencing (TD) [144] based, optical flow (OF) [147] based and motion detector (MD) [153] based. Particularly, for the MD [153] based, we only focus on the extraction of the motion cues

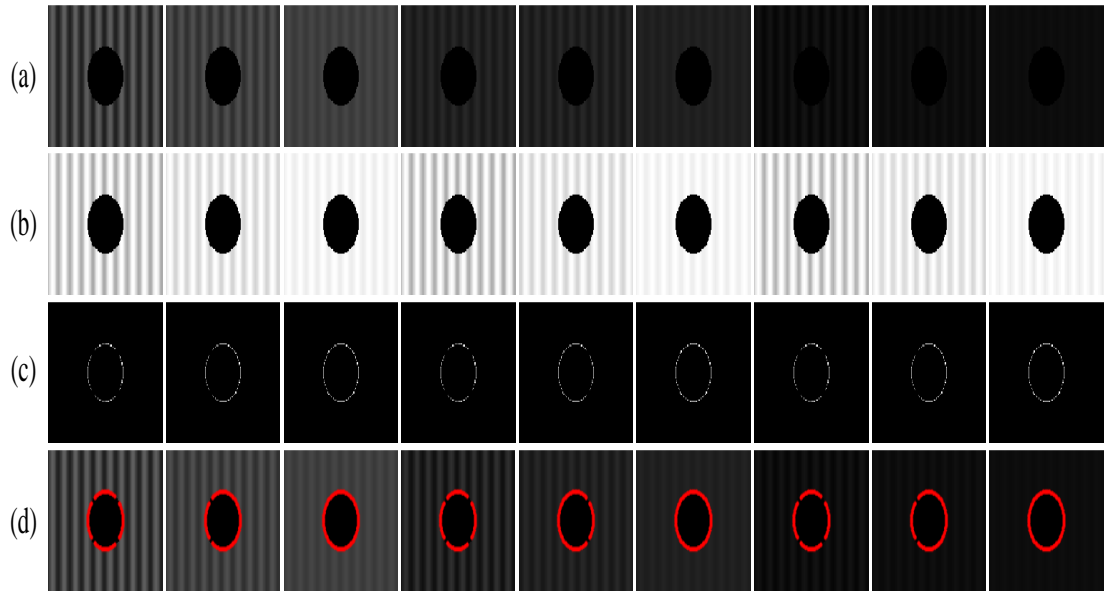


Figure 5.10: From top to bottom row: (a) Input images. They are sample frames from the nine dim-light image sequences, which contain a looming black disk under backgrounds over a range of average luminance ($13 \sim 63$) and RMS contrast ($0.07 \sim 0.35$). (b) Enhanced images. The dim-light images are firstly enhanced by the dark adaptation. (c) The enhanced motion cues. Motion cues are further extracted and enhanced by the spatio-temporal constraint and the neural summation. (d) Detected motion cues (moving edges). The enhanced motion cues in (c) are overlaid on the input images (a), which are visualized by red dots. Motion cues of the looming black disk from different dim-light levels of backgrounds can be effectively extracted by the proposed DLMCE model.

without displaying the neural responses with respect to the directionally selective neurons. In addition, the combined conventional low-light image enhancement methods are CLAHE [19], LIME [20] and Tanaka's [21] that have been introduced in Section 5.2.2.

For clear illustration, pipelines of the implementation of comparative experiments are shown in Fig.5.11. Four processing pipelines are presented: 1) the ground truth of motion cues is obtained by extracting motion trajectory from the raw input that is captured under normal illumination; 2) motion cues output is obtained directly by the motion cues detection method from the raw input that is captured under dim/very dim illumination; 3) motion cues output is obtained by the motion cues detection method from the enhanced input that is produced by the low-light image enhancement method; 4) motion cues output is obtained by the proposed DLMCE model from the raw input under dim/very dim illumination. More details on the comparison are given below.

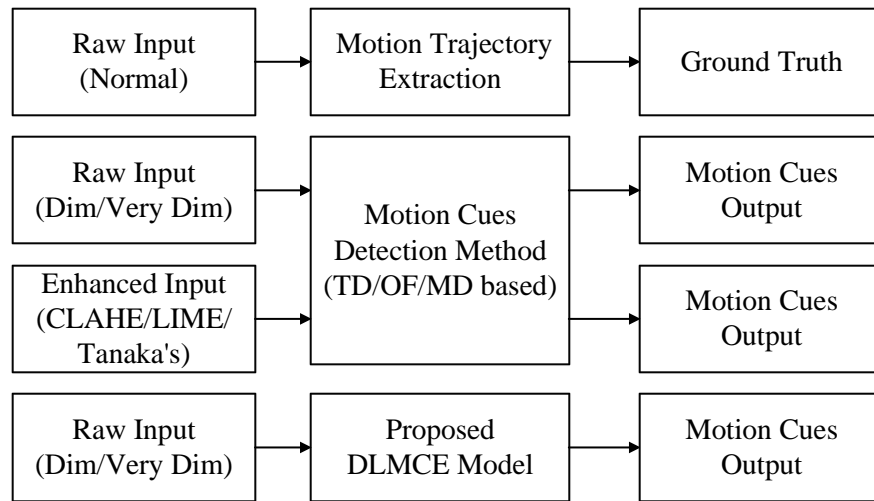


Figure 5.11: Schematic illustration of comparative experiments. From top to bottom: the raw input under normal illumination can be used to extract motion trajectory for obtaining ground truth; the raw input under dark/very dark illumination directly feeds into other motion cues detection methods for obtaining motion cues; the raw input is firstly enhanced by image enhancement methods (CLAHE/LIME/Tanaka's), and then feed into other motion cues detection methods for obtaining motion cues; the raw input feed into the proposed DLMCE model for obtaining motion cues.

Since the ground truth of the motion cue is difficult to define or obtain in low lighting imaging, we adopt two different comparison strategies on synthetic and real datasets. For synthetic datasets, the ground truth is derived from prior knowledge about the motion under normal illumination condition. This is because the synthesized

movement of objects under dim, very dim and normal illumination levels are totally the same, which makes the real motion trajectories of the dim/very dim image sequences capable of extraction. Therefore, the N-cues and F-measure metrics were employed to evaluate the performance between the proposed DLMCE model and the comparative methods on synthetic datasets. However, for real datasets, the ground truth of motion cues are difficult to obtain as the real motion trajectory of a moving object in a dark environment is difficult to label. As a result, we only present the visualization of detected motion cues for an intuitive visual comparison on real datasets. Moreover, we have compared the computational cost between the proposed DLMCE model and comparative methods. Next, comparisons on synthetic and real datasets are described.

1) **Comparison on synthetic dataset:** The synthetic dataset is divided into normal, dim and very dim groups of image sequences. Each group contains translating sequences and looming ones. Sample frames for these sequences are shown in Fig. 5.12. The comparative experiments are conducted as illustrated in Fig. 5.11. First, we extracted the ground truth from the normal group; then, we tested the dim and very dim groups; finally, we show the visualization outputs of motion cues result from the comparative methods (see Fig. 5.13). Note that the ground truth of motion cues are defined as the real motion trajectory in five frame length interval under the frame per second at 120.

Fig. 5.13 shows the visualization of motion cues detection results for the first very dim synthetic image sequences in Fig. 5.12. The evaluation metrics for sixteen tested synthetic sequences are presented in Table 5.2. As can be seen from Fig. 5.13 and Table 5.2, the proposed DLMCE model achieves much better performance than comparative methods. The TD, OF and MD methods show very poor results when the dim/very dim inputs are without any enhancement. Even though the inputs are enhanced by conventional image enhancement methods, their performance is still unsatisfactory. Additionally, the comparison of the computational cost between the proposed DLMCE model and the motion cues detection methods combined with conventional low-light image enhanced methods for synthetic datasets is shown in Table 5.3. Note that the running time is an averaged value for the motion cues detection method combined with

the three enhancement methods over the sixteen sequences. As can be seen, the proposed DLMCE model needs to be in the region of 0.0116s to process a frame, which is much more efficient than the comparative methods.

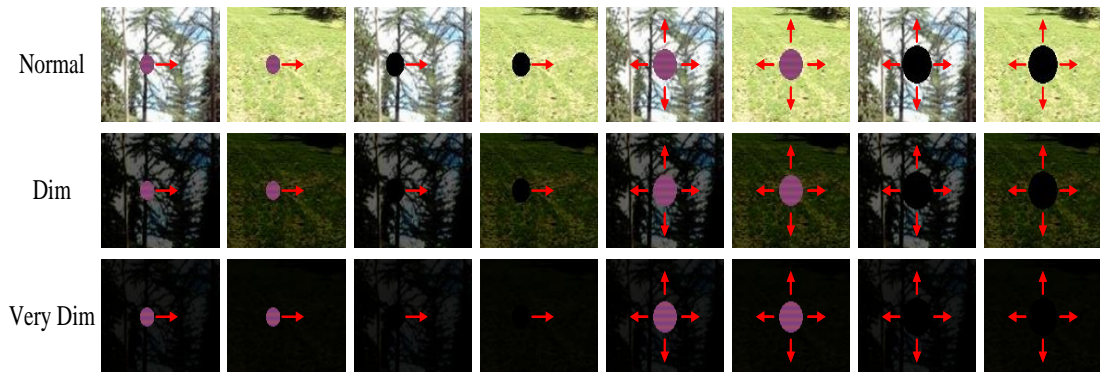


Figure 5.12: Sample frames from synthetic datasets, including translating and looming objects under normal, dim and very dim illumination of backgrounds. Specifically, the degree of the background's darkness is controlled by the Matlab gamma intensity correction function, and the maximum output intensity is limited to the range of [12, 102]. From top to bottom rows: a translatory/looming colorful textured disk and a translatory/looming black one are shown in the sample frames. Note that the normal illumination dataset is used for extracting ground truth. Dim and very dim datasets as inputs feed into motion cues detection methods for conducting comparative experiments.

2) **Comparison on real dataset:** The real dataset contains six image sequences which were captured by a GoPro camera under dim-light environments with artificial lights or without ones. From sequences one to four, each displays a translatory/looming ball under dim/very dim (with/without artificial light) indoor environments. Sequences five and six are captured at nighttime in outdoor environments, including a translatory and an approaching person separately. Note that five sample frames of each tested sequence have been shown in comparative experiments.

We compared the visual performance between the proposed DLMCE model and the comparative methods on the six real image sequences. The visualization results are shown in Fig. 5.14–Fig. 5.17, where red dots indicate the detected motion cues. As can be seen, for TD, OF and MD methods, motion cues are hardly detected when there is no enhancement for the dim-light input. And the extracted motion cues are not enough or contain a lot of noise, even with the enhanced input by the conventional enhancement methods. This is because the motion cue is related to changes in luminance and the effect of simply enhancing low-light input is limited. The results in Fig. 5.14–Fig. 5.17

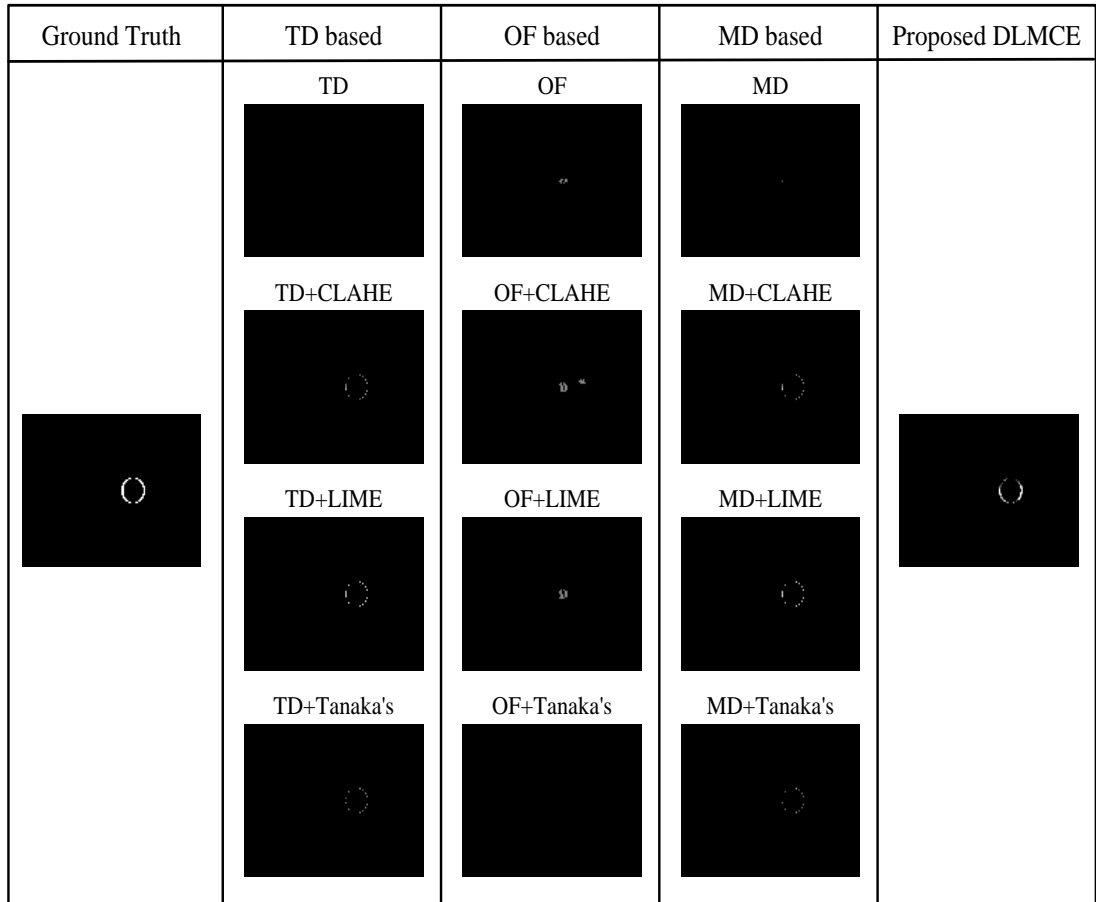


Figure 5.13: Results of motion cues detection from the first sequence of the very dim group (see Fig. 5.12), which show a detailed evaluation between the proposed DLMCE model and the comparative methods.

Table 5.2: N-CUES/F-MEASURE COMPARISON. THE VALUE IS AVERAGED OVER THE SIXTEEN SYNTHETIC DIM-LIGHT IMAGE SEQUENCES. **THE LARGER THE N-CUES AND THE F-MEASURE, THE MORE MOTION CUES CAN BE ACCURATELY DETECTED.**

Methods	Avg. N-cues	Avg. F-measure (%)
TD	0	0
TD+CLAHE	7.4	19.28
TD+LIME	11.0	28.21
TD+Tanaka's	11.1	26.69
OF	0.8	1.65
OF+CLAHE	3.4	3.79
OF+LIME	3.4	6.39
OF+Tanaka's	3.0	5.84
MD	0.3	1.13
MD+CLAHE	6.7	17.76
MD+LIME	10.3	26.85
MD+Tanaka's	10.1	24.40
Proposed DLMCE	43.0	67.67

Table 5.3: RUNNING TIME COMPARISON. THE VALUE IS AVERAGED OVER THE SIXTEEN SYNTHETIC DIM-LIGHT IMAGE SEQUENCES WITH 160 FRAMES WHILE WITH FRAME SIZE OF 128 PIXELS (HORIZONTAL) \times 96 PIXELS (VERTICAL).

Methods	Avg. Time (s/frame)
TD+CLAHE/LIME/Tanaka's	0.0196
OF+CLAHE/LIME/Tanaka's	0.3572
MD+CLAHE/LIME/Tanaka's	0.0158
Proposed DLMCE	0.0116

show the superior performance of the proposed DLMCE model in enhancing dim-light motion cues over the comparative methods. Table 5.4 presents the comparison of the computational costs of different methods in processing the real dataset, which also demonstrates that the proposed DLMCE model is much more efficient than the comparative methods.

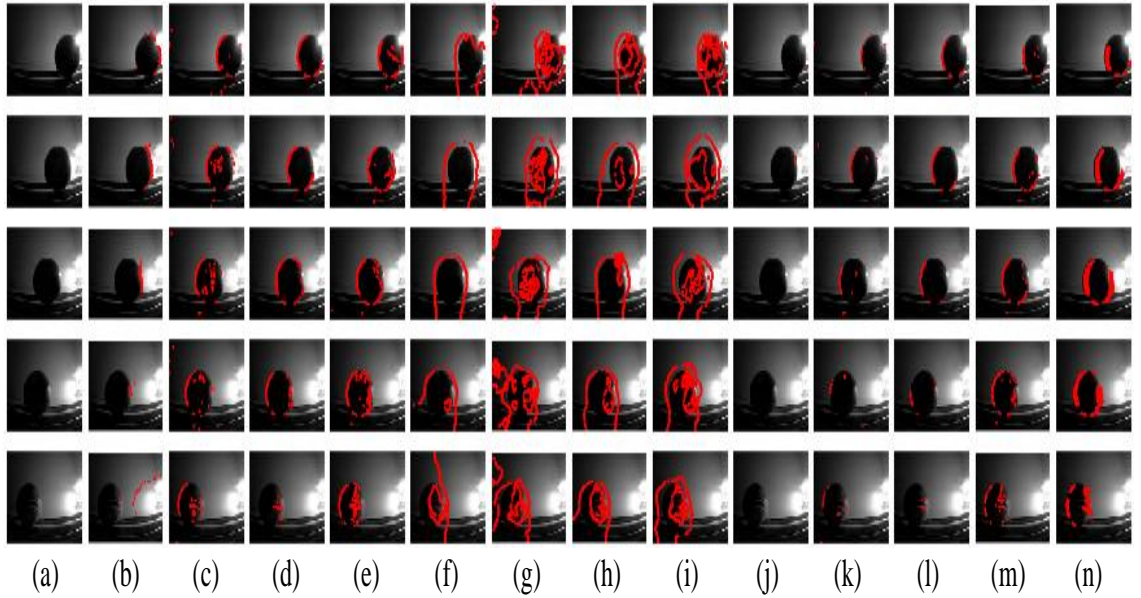


Figure 5.14: Comparison results of detected translating motion cues overlaid on the dim inputs between the proposed DLMCE model and other methods (**tested sequence one, a translatory black ball in a dim environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE.

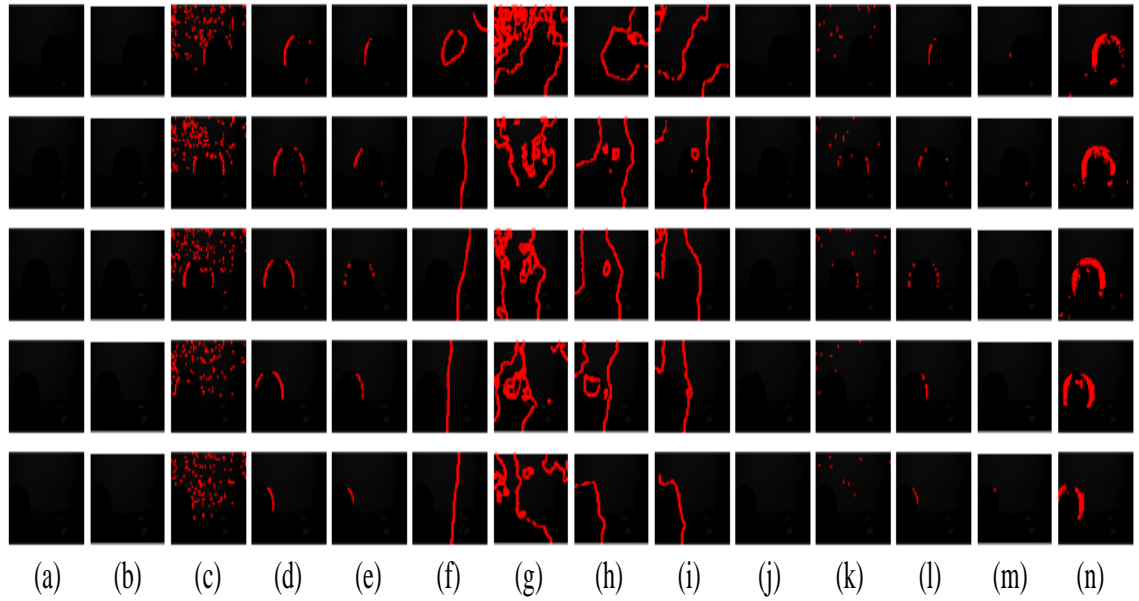


Figure 5.15: Comparison results of detected translating motion cues overlaid on the very dim inputs between the proposed DLMCE model and other methods (**tested sequence two, a translatory black ball in a very dim environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE.

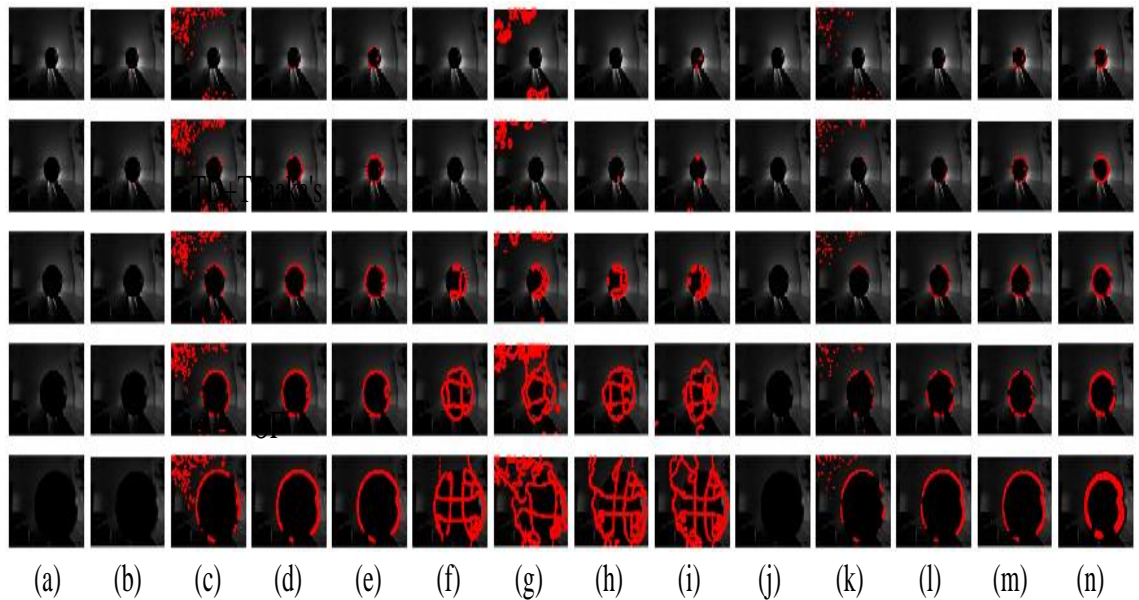


Figure 5.16: Comparison results of detected looming motion cues overlaid on the dim inputs between the proposed DLMCE model and other methods (**tested sequence three, a looming black ball in a dim indoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE.

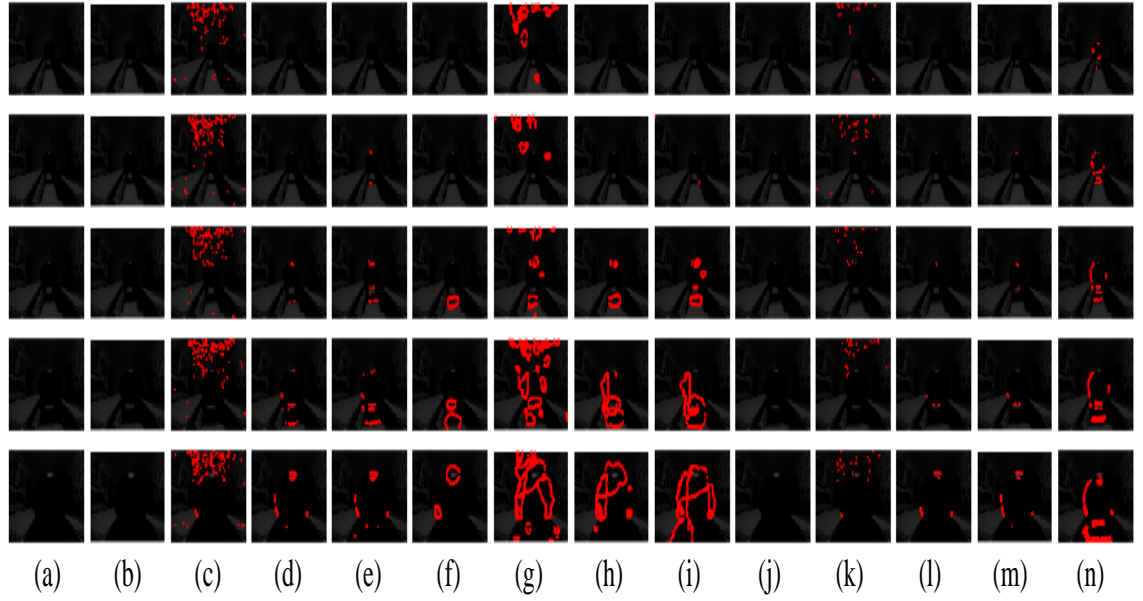


Figure 5.17: Comparison results of detected looming motion cues overlaid on the very dim inputs between the proposed DLMCE model and other methods (**tested sequence four, a looming ball in a very dim indoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE.

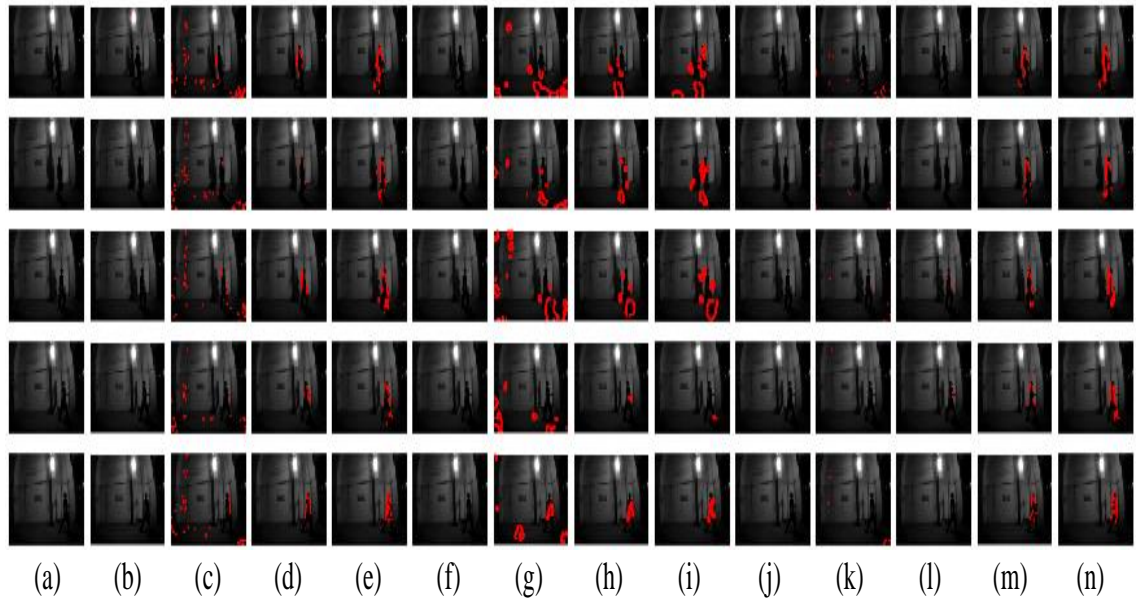


Figure 5.18: Comparison results of detected translating motion cues overlaid on the dim inputs between the proposed DLMCE model and other methods (**tested sequence five, a translatory person in a dim outdoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE.

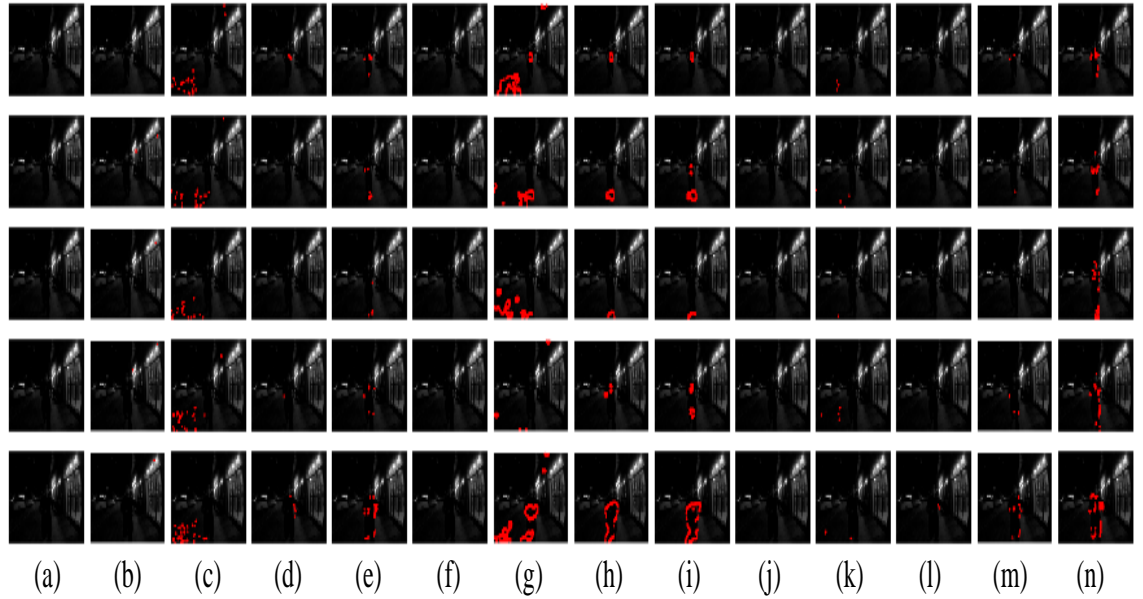


Figure 5.19: Comparison results of detected looming motion cues overlaid on the very dim inputs between the proposed DLMCE model and other methods (**tested sequence six, a looming person in a very dim outdoor environment**), which are visualized by red dots. From top to bottom row are frame10 to frame50 respectively. (a) Input frames. (b) TD. (c) TD+CLAHE. (d) TD+LIME. (e) TD+Tanaka's. (f) OF. (g) OF+CLAHE. (h) OF+LIME. (i) OF+Tanaka's. (j) MD. (k) MD+CLAHE. (l) MD+LIME. (m) MD+Tanaka's. (n) Proposed DLMCE.

Table 5.4: RUNNING TIME COMPARISON. THE VALUE IS AVERAGED OVER THE SIX REAL DIM-LIGHT IMAGE SEQUENCES WITH 300 FRAMES WHILE WITH FRAME SIZE OF 128 PIXELS (HORIZONTAL) \times 96 PIXELS (VERTICAL).

Methods	Avg. Time (s/frame)
TD+CLAHE/LIME/Tanaka's	0.0263
OF+CLAHE/LIME/Tanaka's	0.2405
MD+CLAHE/LIME/Tanaka's	0.0255
Proposed DLMCE	0.0115

5.3 Chapter Summary

In this chapter, we have proposed a bio-inspired DLMCE model for dim-light motion cues enhancement in low lighting conditions. The key of the model is the combination of canonical neural computation and neural summation. Inspired by nocturnal animals' visual mechanisms, the DLMCE neural network model integrates the processings of dark-adaptation, spatio-temporal constraint and neural summation in spatial and temporal domains for faint motion cues extraction. The dark-adaptation processing is able to increase image intensity and contrast. The spatio-temporal constraint processing is responsible for extracting prominent motion cues. The neural summation processing in spatial and temporal domains can further enhance the extracted motion cues. Comprehensive experiments on both the synthetic and real datasets have demonstrated that the proposed DLMCE model is effective and efficient in enhancing faint motion cues, and performed superiority over the comparative methods.

In the future, we will further investigate the potential applications of the DLMCE model integrated into other directional selective neural network models to handle low-light visual scenes, e.g. night navigation, dim-light collision detection, etc.

Chapter 6

Bio-inspired Collision Detection in Dim Light Environments

The ability to detect imminent collisions is essential for visual navigation under low light conditions. Autonomous vehicles or robots need to avoid colliding objects without human intervention when they run in a complex and dim-light environment. To recognize frontal obstacles, extracting looming cues is vital. However, it is rather challenging for these autonomous systems to detect collisions reliably due to the low signal-to-noise ratio (SNR). As a result, not only is the recognition of approaching objects important, but also the enhancement of looming cues is necessary for collision detection at low light intensities.

The lobula giant movement detector LGMD1 is an identified visual neuron in the optic lobe of the locust, which shows a strong preference for approaching objects [36], [193]. Since the LGMD1 can protect the locust from collision by recognizing objects moving directly toward it, the collision rates are low despite their swarm migration with thousands of individuals. Numerous biological researches [1], [24], [33], [62], [71], [194] aim to investigate the collision selectivity of the LGMD1 neuron and, at the same time, a lot of LGMD1 based neural network models also have been proposed for collision detection applications [5], [7]–[9], [31]. Although these LGMD1 based models are able to show their effectiveness in detecting looming objects under certain conditions, they cannot cope with very dim light situations where motion cues are buried in the dark.

It is known that nocturnal insects show remarkable visual abilities at low levels of ambient light due to their specific visual processing strategies. Neural mechanisms of nocturnal vision concerning how visual signals are processed are the focus of intense research [10], [17], [195], [196]. Inspired by the neural mechanisms within nocturnal animals, methods have been proposed for improving artificial vision systems' visual performance in low light conditions. For example, spatial and temporal summation strategies have been adopted to reduce significant noise arising from low light image enhancement and thus improve the visual quality [11], [42]. For motion detection, we have proposed a dim-light motion cues enhancement neural network model (DLMCE) that has been introduced in Chapter 5. The key of the DLMCE model is that it only enhances faint motion cues by a series of canonical neural computations and neural summation strategies. Experiments have demonstrated the effectiveness of the DLMCE in Chapter 5. Therefore, the proposed DLMCE model should be possible to combine with directional selective neural network models for motion detection applications in low light environments, such as the LGMD1 based dim-light collision detection.

Existing LGMD1 models can be generally divided into two categories in terms of visual signal processing. In the first category [5], [7], [31], [197], both dark and light looming cues (expanding edges) are processed together in the model. In the second category [8], [9], [156], motion cues are separated into positive (ON) and negative (OFF) contrast polarity, which are processed in ON and OFF pathways separately. Since the previous biological findings indicate that the behavior of ON and OFF followed parallel courses was not found at low luminance levels [161], the first category LGMD1 models are given priority to develop the bio-inspired dim-light collision detection model. The classical four-layered model [5] has the simplest structure compared with other variants of the first category LGMD1. We therefore combine it with the DLMCE model to achieve dim-light collision detection with less computational cost.

Inspired by insect vision, we propose a dim-light collision detection model that combines the DLMCE model and the classical four-layered LGMD1 model for detecting looming objects under low light conditions. The DLMCE model is responsible for enhancing motion cues. Moreover, the enhanced motion cues as inputs are passed to

the first layer of the four-layered LGMD1 model for achieving looming sensitivity. The main contribution of this work is the integration of the DLMCE and the LGMD1 so that the proposed model can recognize and respond to extremely weak looming stimuli in dim-light environments.

The remainder of this chapter is organized as follows. Section 6.1 introduces the formulation of the proposed bio-inspired dim-light collision detection model. Section 6.2 reports experimental results, including the effectiveness of looming motion detection and the performance comparisons with comparative LGMD1 models on the real dim-light dataset. Finally, we conclude this chapter in Section 6.3.

6.1 Formulation of the Bio-inspired Dim-light Collision Detection Model

The proposed bio-inspired dim-light collision detection model is shown in Fig. 6.1. The model combines the DLMCE model with the classical four-layered LGMD1 model [5], [22] for detecting looming objects in dim light environments. It utilizes the DLMCE model to enhance dim-light motion cues, and then these enhanced motion cues are fed into the first layer of the LGMD1 model. In the following subsections, we have elaborated on its components.

6.1.1 DLMCE layer

The DLMCE layer is composed of six processing sublayers (L, P, E, I, S and M). Firstly, the L sublayer implements the dark adaptation processing with a series of canonical neural computations. Then, the P sublayer computes the changes in luminance to obtain potential motion cues and uses the spatio-temporal constraint strategy to extract prominent motion cues. Finally, these extracted motion cues are passed to the E, I, S and M sublayers, which are further enhanced by the neural summation strategy in temporal and spatial domains.

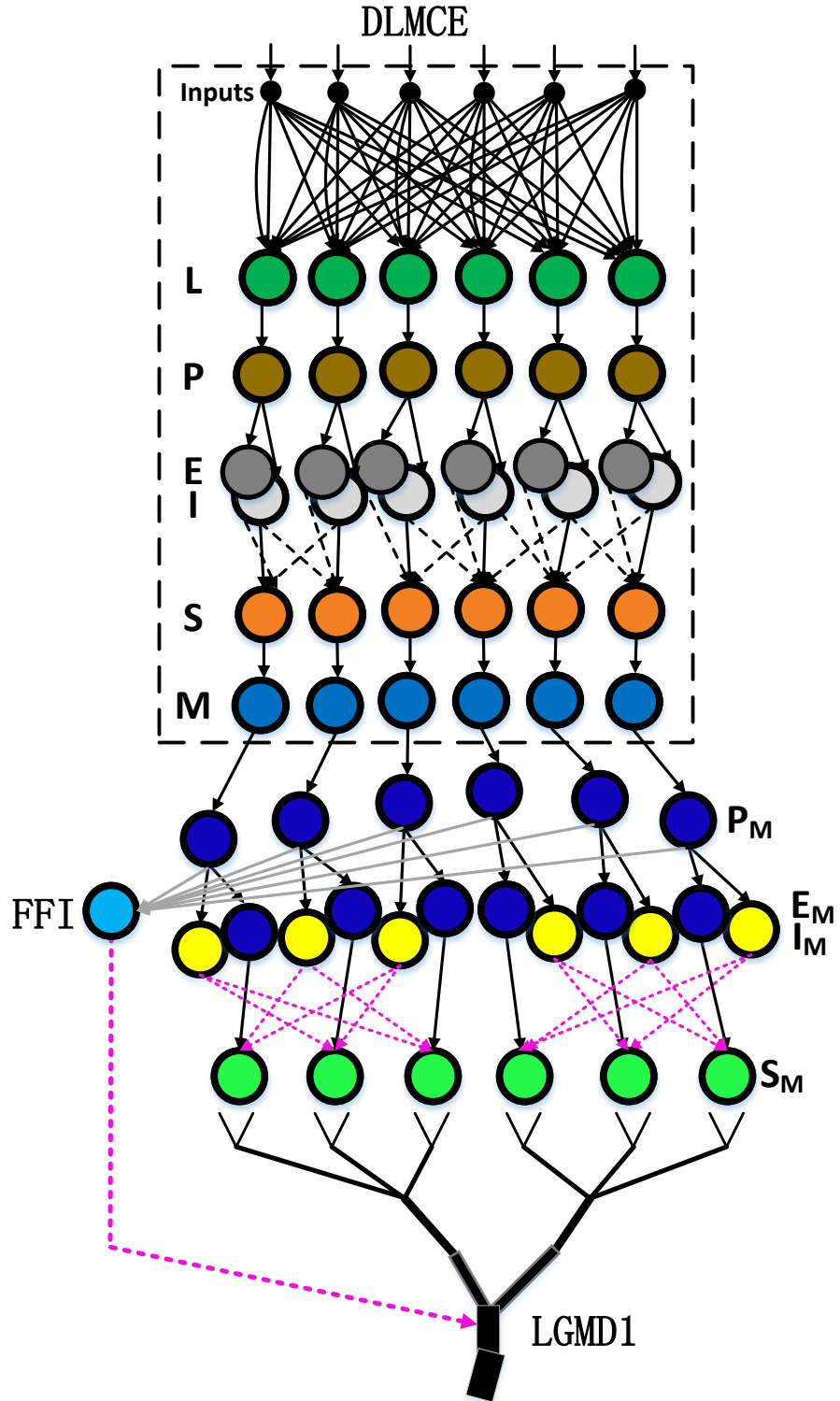


Figure 6.1: The proposed bio-inspired dim-light collision detection model is composed of five layers (DLMCE, P_M , E_M , I_M , S_M), and the LGMD1 cell and the feedforward inhibition (FFI) cell. Specifically, it combines the DLMCE model introduced in Chapter 5 with the classical four-layered LGMD1 model [5], [22] together where the outputs of DLMCE feed the first layer of the LGMD1.

6.1.1.1 L sublayer

The dark adaptation processing in the L sublayer involves the power law adaptation [108], the divisive normalization [15], [115] and the adaptive rescaling operation [181], [190]. More details of these canonical neural computations are described as follows. First, the operation of the power law for image intensity is mathematically defined by,

$$S_L(x, y, t) = I^n(x, y, t) \quad (6.1)$$

where I denotes the value of image intensity and the exponent n is the adaptation parameter. Note that the value of n is set lin accordance with Chapter 5. Then, the perceived power-law image intensity is normalized by,

$$S_L'(x, y, t) = \frac{S_L(x, y, t)}{S_L(x, y, t) + S_I_m(t)} \quad (6.2)$$

where S_I_m indicates the average intensity of the input image. It is computed by the following equation:

$$S_I_m(t) = \frac{1}{N} \sum_{x=1}^r \sum_{y=1}^c I(x, y, t) \quad (6.3)$$

where N denotes the number of pixels, r and c represent the number of image rows and columns respectively. Note that the value of N equals $r * c$. Finally, the normalization intensity is rescaled by,

$$S_L_o(x, y, t) = \frac{S_L'(x, y, t) - S_L'_{min}}{S_L'_{max} - S_L'_{min}} \quad (6.4)$$

where S_L_o represents the rescaled neural response that can maximize information transmission. $S_L'_{max}$ and $S_L'_{min}$ indicate the maximum and the minimum value in S_L' separately.

6.1.1.2 P sublayer

The P sublayer is responsible for extracting prominent motion cues. It calculates the changes in luminance. Note that the luminance is filtered by a temporal low-pass

filter. That is,

$$S_P(x, y, t) = \text{abs} \left(S_L_o(x, y, t) - S_L_o(x, y, t - \tau) \right) \quad (6.5)$$

$$S_L_o(x, y, t) = \frac{1}{\int_{t-T}^t S_L_o(i) di} \cdot \int_{t-T}^t S_L_o(T - i) \cdot S_L_o(x, y, i) di \quad (6.6)$$

$$S_L_o(t) = e^{-t/\sigma_t} \quad (6.7)$$

$$S_T = n_f * \tau \quad (6.8)$$

where $S_P(x, y, t)$ represents the luminance increments or decrements which can be regarded as potential motion cues, and S_L_o denotes the filtered luminance. Note that σ_t indicates the decay time constant, n_f is the number of frames, and τ is the frame interval.

For obtaining prominent motion cues, the processing of spatio-temporal constraint is also implemented in the P sublayer. The changes in luminance are firstly constrained in the temporal domain. More concretely, the luminance changes between time t and $t - \Delta t$ are considered as motion cues when they are greater than a specific proportion of the average luminance level at the current time. The temporal constraint is achieved by an adaptive thresholding processing. That is,

$$S_P(x, y, t) = \begin{cases} S_P(x, y, t) & \text{if } S_P(x, y, t) \geq T_p \\ 0 & \text{otherwise} \end{cases} \quad (6.9)$$

where T_p is the threshold that is determined by the mean value of S_L_o , defined by

$$S_T_p(t) = \alpha \cdot \frac{\sum_{x=1}^r \sum_{y=1}^c S_L_o(x, y, t)}{N} \quad (6.10)$$

where the coefficient α can be adjusted. Note that the value of α affects the T_p so that it can further influence the extraction of prominent motion cues. Then, the spatial constraint is carried out by the normalized scaling processing, as described by the following equation,

$$S_P_o(x, y, t) = \frac{S_P(x, y, t)}{S_P_{max}} \quad (6.11)$$

where the output of the P cell is constrained in the spatial domain, and \tilde{P}_{max} indicates the maximum value of \tilde{P} in the spatial domain at time instant t .

6.1.1.3 E/I sublayer

The luminance changes are outputs of the P cells, which are collected in a specific time interval m by the E/I cells. As a result, the outputs of E/I cells are actually averaged signals by the neural summation in the temporal domain. The final output of the E cell is given by,

$$S_E(x, y, t) = \frac{1}{m} \cdot \int_{t-m}^t S_P_o(x, y, i) di \quad (6.12)$$

Note that the I cell has the same output as the E cell.

6.1.1.4 S sublayer

The S cell accumulates the excitation, the delayed self-inhibition at the same retinotopic position and its neighboring delayed excitations. The spatio-temporal summation process can be mathematically described by,

$$S_E_N(x, y, t - k\tau) = \sum_{i=-1}^1 \sum_{j=-1}^1 S_E(x + i, y + j, t - k\tau) S_w_e(i, j) \quad (6.13)$$

$$S_w_e = \frac{1}{8} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (6.14)$$

$$S_S_{k\tau}(x, y, t) = S_E(x, y, t) + S_E_N(x, y, t - k\tau) - S_I(x, y, t - k\tau) \quad (6.15)$$

$$S_S_{k\tau}(x, y, t) = \begin{cases} 1 & \text{if } S_{k\tau}(x, y, t) \geq T_s, k \in Z^+ \\ 0 & \text{otherwise} \end{cases} \quad (6.16)$$

where $S_E_N(x, y, t - k\tau)$ denotes the excitations which are passed from the neighboring pixels of position (x, y) with delays of k frame intervals. S_w_e is the weight matrix

for the neighboring excitations. $S_I(x, y, t - k\tau)$ indicates the inhibition passed from the same spatial position with the same time delays. The $S_{k\tau}$ represents enhanced motion cues received by the S cell. If the accumulated value in the S cell exceeds the threshold T_s , then a spike is generated.

6.1.1.5 M sublayer

The M sublayer implements a temporal summation where the M cell collects motion signals transmitted with different time delays. The temporal summation process is given by:

$$S_M(x, y, t) = \frac{1}{3} \cdot \sum_{k=1}^3 S_{k\tau}(x, y, t) \quad (6.17)$$

where S_M indicates the averaged motion signal within a time duration of 3τ .

6.1.2 P_M layer

The P_M cell receives the enhanced motion cues passed from the M sublayer of the DLMCE layer. Since the output of the DLMCE layer as inputs feed the LGMD1 model, the P_M layer receives the enhanced motion cues from the M sublayer. Note that the output value is multiplied by 255 after passing to the P_M cell. That is,

$$P_M(x, y, t) = 255 * S_M(x, y, t) \quad (6.18)$$

where P_M denotes the enhanced luminance changes that correspond to the range of $0 \sim 255$.

6.1.3 E_M/I_M layer

In this layer, two separate groups of cells are formed by the output of the P_M cells. One is the excitatory cells, which pass the excitation directly to their retinotopic counterpart in the S_M layer. The excitation in an E_M cell is the same as that in a P_M cell. The other is the lateral inhibition cells, through which inhibition is passed to their retinotopic counterpart's neighboring cells in the S layer with one frame interval delay. The

gathered strength of inhibition in a I_M cell is given by,

$$I_M(x, y, t) = \sum_i \sum_j P_M(x + i, y + j, t - \tau) w_I(i, j), \quad (if \ i = j, \ j \neq 0) \quad (6.19)$$

$$w_I = \begin{bmatrix} 0.125 & 0.25 & 0.125 \\ 0.25 & 0 & 0.25 \\ 0.125 & 0.25 & 0.125 \end{bmatrix} \quad (6.20)$$

where I_M indicates the calculated lateral inhibition, $w_I(i, j)$ is the local inhibition weight matrix, and the value of τ equals one frame interval.

6.1.4 S_M layer

The excitations from the E_M cells and inhibitions from the I_M cells are summed by the S_M cells using the following equation:

$$S_M(x, y, t) = E_M(x, y, t) - I_M(x, y, t) W_I \quad (6.21)$$

where W_I is the inhibition weight.

6.1.5 LGMD1 cell

The membrane potential of the LGMD1 cell at time instant t is summed after the S layer with a rectifying operation. The process can be described by the following equation,

$$MP(t) = \sum_x \sum_y \text{abs}(S_M(x, y, t)) \quad (6.22)$$

Then, the membrane potential of the LGMD1 cell is normalized. That is,

$$NMP(t) = 1 - (1/\exp(\beta^{-1} \cdot MP(t) \cdot n_{\text{cell}}^{-1})) \quad (6.23)$$

where the value of NMP is in the range of $(0 \sim 1)$, and parameters β and n_{cell} are set in accordance with [156].

6.1.6 Spiking Mechanism

The spiking mechanism in the LGMD1 neuron can be described by the following equations,

$$Spike(t) = \begin{cases} 1 & \text{if } NMP(t) \geq T_{thresh} \\ 0 & \text{otherwise} \end{cases} \quad (6.24)$$

where the spike is produced when the value of NMP exceeds the threshold T_{thresh} . Moreover, the alarm time (A_T) is given by,

$$A_T = \begin{cases} t & \text{if } \int_{t-t_n}^t Spike(i) \geq n_{sp} \\ 0 & \text{otherwise} \end{cases} \quad (6.25)$$

where the value of A_T does not equal ZERO indicates that there are at least n_{sp} successive spikes in the length of time t_n and a collision alarm will be set. In addition, the alarm time can be discretized into the alarm frame. Note that the value of T_{thresh} and n_{sp} are set in accordance with [7].

6.1.7 The Feed Forward Inhibition (FFI)

The feedforward inhibition (FFI) is used to shut down the spiking of LGMD1 when substantial changes in luminance occur in the visual scene caused by turning or changes in ambient lighting conditions. The FFI signal is gathered from P cells with one frame time delay τ , which is mathematically described by,

$$FFI(t) = \sum_x \sum_y P_M(x, y, t - \tau) \cdot n_{cell}^{-1} \quad (6.26)$$

where n_{cell} indicates the number of cells. If the value of FFI exceeds its threshold T_{FFI} , spikes produced by the LGMD1 will be inhibited immediately. Note that the value of T_{FFI} is set to 10 in accordance with [7].

6.1.8 Parameters of the System

The parameters of the proposed bio-inspired dim-light collision detection model are listed in Table 6.1 based on current trials in this chapter. These parameters are determined according to Chapter 3 and Chapter 5. In the following experiments, they are kept unchanged unless stated.

Table 6.1: PARAMETERS OF THE PROPOSED BIO-INSPIRED DIM-LIGHT COLLISION DETECTION MODEL

Eq.	Parameters	Eq.	Parameters
(6.1)	$n = 0.4$	(6.21)	$W_I = 0.3$
(6.7)	$\sigma_t = 4$	(6.23)	$\beta = 0.25, n_{cell} = 12288$
(6.8)	$n_f = 3, \tau = 8.33 \text{ ms}$	(6.24)	$T_{thresh} = 0.7$
(6.10)	$\alpha = 0.2$	(6.25)	$n_{sp} = 4$
(6.16)	$T_s = 0.35$		

6.2 Experimental Results

6.2.1 Experimental Setup

1) Data Set: The proposed bio-inspired dim-light collision detection model is evaluated on a real dim-light looming dataset. The dataset consists of four dim-light looming video clips that are captured by a GoPro camera in indoor and outdoor environments, including 800 image frames. These images have a resolution of 960×1280 pixels at 120 fps. Note that the image was resized to 96×128 pixels using the ‘imresize’ function in the Matlab image processing toolbox before being feed to the neural network model. The dim-light looming dataset is shown in Fig. 6.2 that contains sixteen sample frames from the four tested image sequences (video clips).

2) Evaluation Criteria: We employ the normalized membrane potential (NMP) curve to show the output of LGMD1 models. If the result of an LGMD1 responding to visual stimuli is correct, it indicates a successful detection. The result is related to the alarm frame (A_F).

2) Implementation: In the research into the LGMD1 based models, the proposed models are devoted to shaping the collision selectivity. However, these models cannot

work well in dim light environments due to the extremely low luminance and contrast. In this chapter, we have used four real dim light image sequences with different backgrounds to test the effectiveness of the proposed model. We have also compared the performance between the proposed model and the two existing LGMD1 models when challenged with the low-light image sequences.

6.2.2 Effectiveness of Looming Motion Detection

To verify the effectiveness of the proposed model in the detection of looming motion, we display enhancement results for low-light images and faint motion cues as well as the response of LGMD1 neuron, as shown in Fig. 6.3 and Fig. 6.4. Specifically, Fig. 6.3 shows the results for looming balls in indoor environments (tested sequences one and two), whilst Fig. 6.4 shows the results for approaching persons in outdoor environments (tested sequences three and four). Among these results, the enhanced intensity image and motion cues are outputs of the L sublayer and the M sublayer in the DLMCE layer respectively. Moreover, the response of the LGMD1 neuron is shown with the curve of normalized membrane potential (NMP). It is worthy to note that all displayed images correspond to the last frame (frame 200), including the dark image as the input.

As can be seen from the enhanced intensity image and motion cues in Fig. 6.3 and Fig. 6.4, the DLMCE layer is effective in raising the intensity of the dark pixel and extracting faint motion cues buried in the dark. Although some moving edges (motion cues) of the looming object are missed, the extracted prominent motion cues are further enhanced by the neural summation strategy in spatial and temporal domains (see the thick edges in Fig. 6.3 and Fig. 6.4). The response curves of the LGMD1 neuron have verified that the proposed model can achieve the detection of the looming object.

6.2.3 Comparison with Comparative LGMD1 Models

To further verify the effectiveness of the proposed model, we compared the performance with two comparative LGMD1 models. The first comparative LGMD1 model is the classical four-layered model proposed by Rind [5] and the other comparative

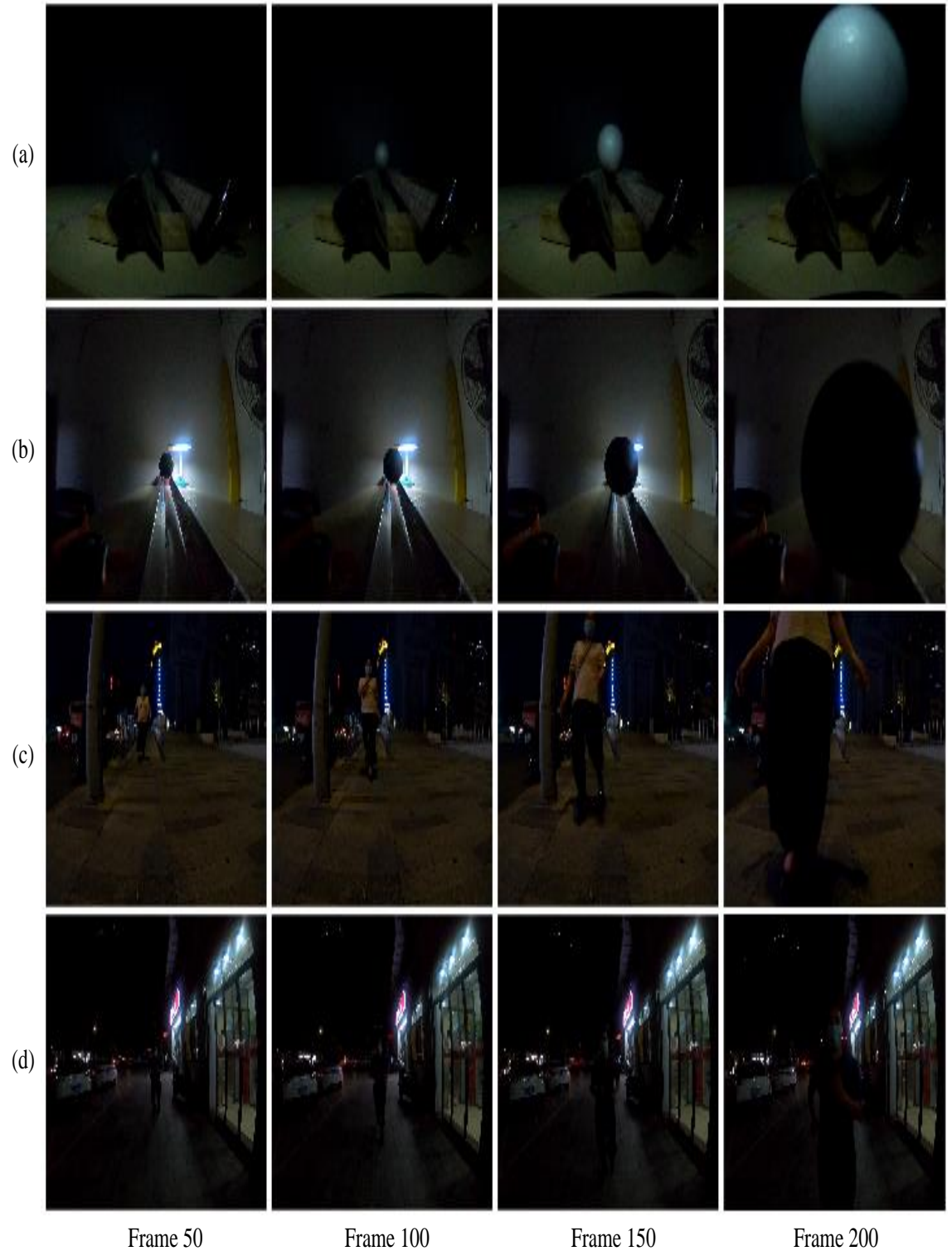


Figure 6.2: Sample frames from the four tested image sequences in the dim-light dataset. From the first column to the fourth column are frame 50, frame 100, frame 150 and frame 200 separately. From top row to bottom row are sequence one to sequence four: (a) sequence one, a looming white ball; (b) sequence two, a looming black ball; (c) sequence three, an approaching woman wearing a light T-shirt; (d) sequence four, an approaching man wearing a dark T-shirt. It is worthy to note that sample frames in (a), (b) and (c), (d) are captured in indoor and outdoor environments individually.

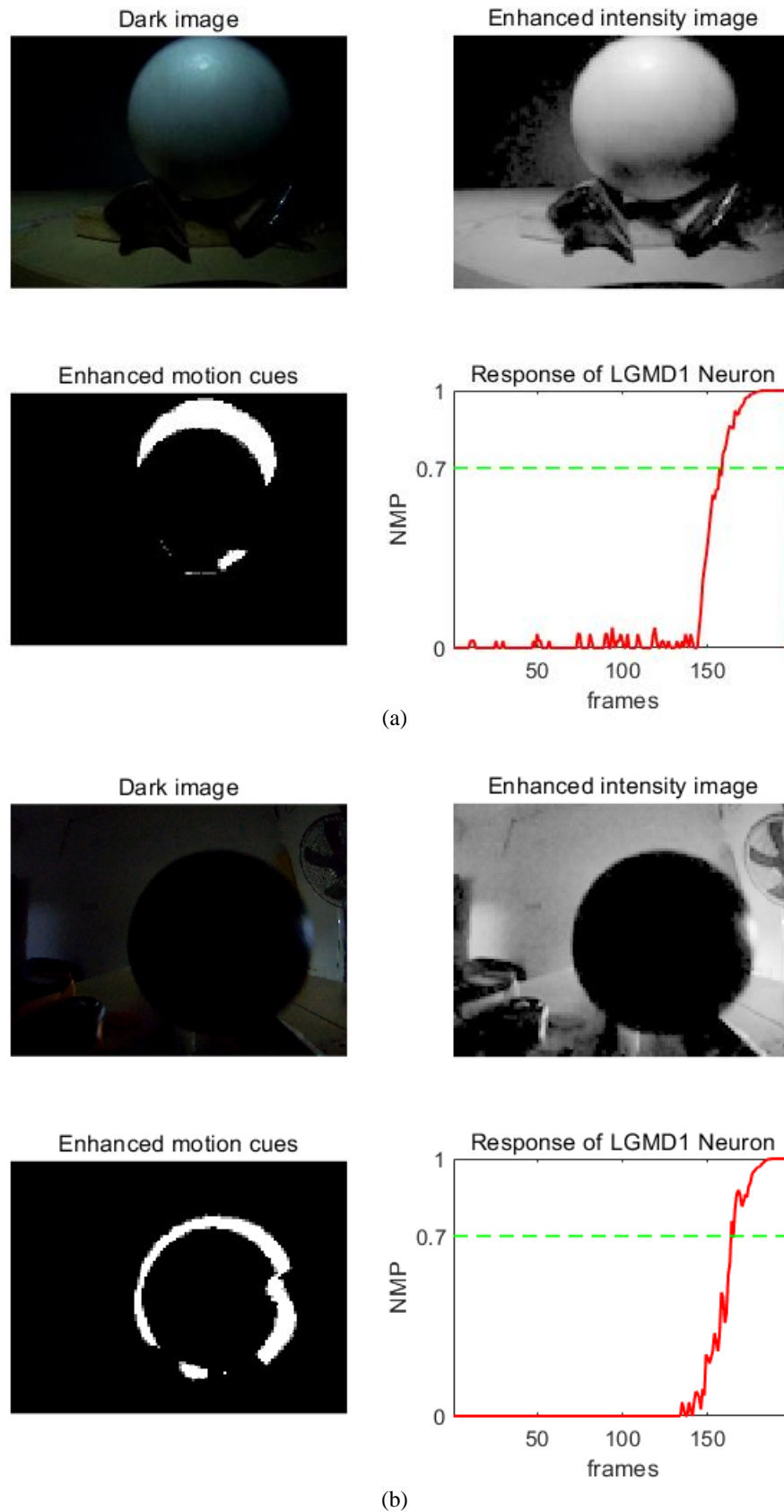


Figure 6.3: Experimental results of the proposed model for sequence one (subfigure (a)) and sequence two (subfigure (b)). In each subfigure, it includes the dark image (input), enhanced intensity image (output of L sublayer in DLMCE layer), enhanced motion cues (output of M sublayer in DLMCE layer) and response of LGMD1 neuron (NMP curve). Note that each exhibited image in each subfigure is the result of frame 200.

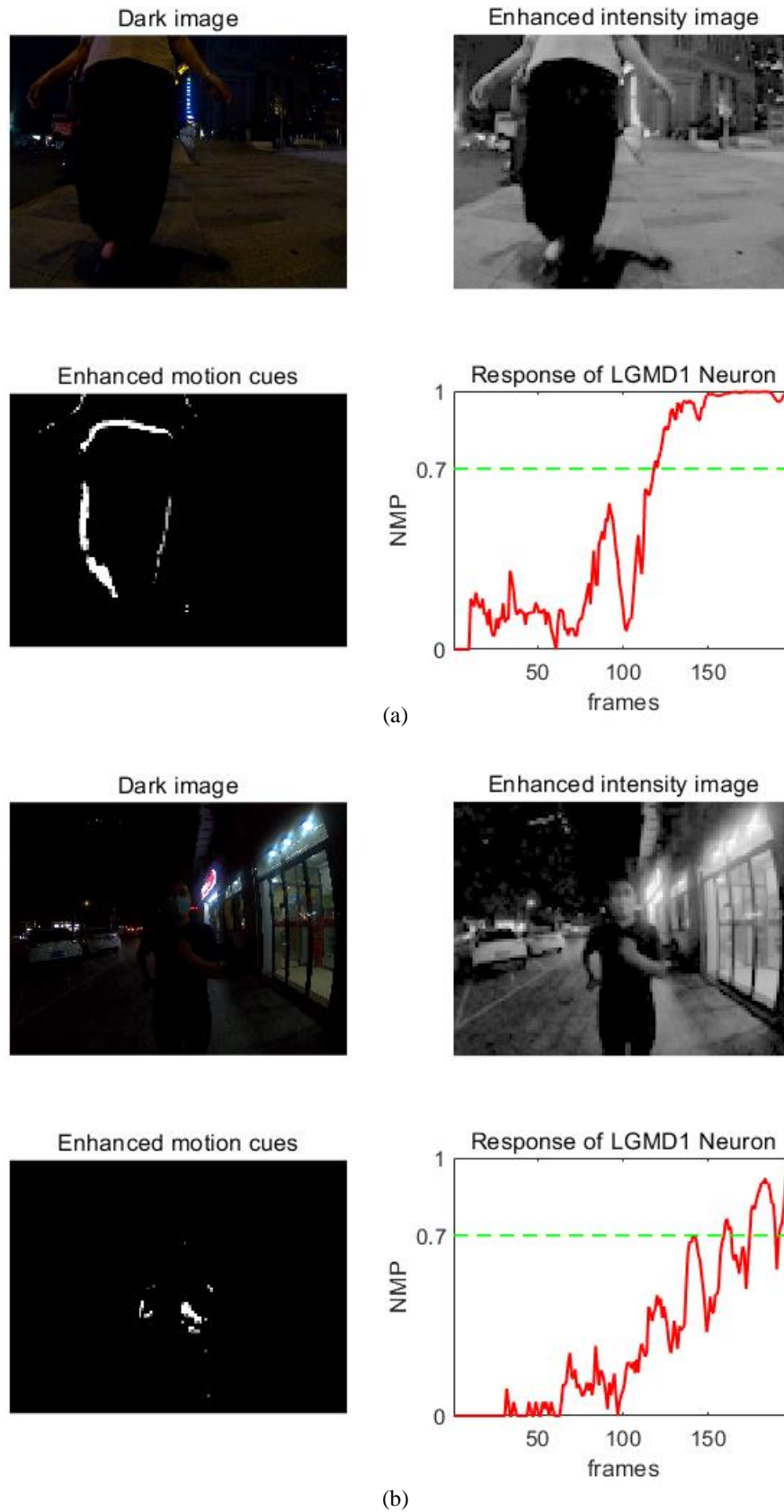


Figure 6.4: Experimental results of the proposed model for sequence one (subfigure (a)) and sequence two (subfigure (b)). In each subfigure, it includes dark image (input), enhanced intensity image (output of L sublayer in DLMCE layer), enhanced motion cues (output of M sublayer in DLMCE layer) and response of LGMD1 neuron (NMP curve). Note that each exhibited image in each subfigure is the result of frame 200.

LGMD1 model is the variant of the first one proposed by Yue [7]. Note that all models are challenged with low-light image sequences directly.

From the NMP curves in Fig. 6.5 and Fig. 6.6, we can observe that Yue's model [7] can only produce little or no responses to looming stimuli when it is challenged with very dark image sequences. Although the responses of Rind's model [5] do not tend to zero, it cannot produce a timely collision alarm for objects approaching on a direct collision. For the proposed model, it can trigger an effective warning of collision and the alarm frame among the four tested sequences occurs at frame 162, frame 169, frame 122 and frame 163 individually. Note that the collision frame occurs after frame 200. This indicates that the proposed model can correctly respond to faint looming stimuli and produce a timely collision alarm, which significantly outperforms the two comparative models.

6.3 Chapter Summary

In this chapter, we have proposed a bio-inspired dim-light collision detection model for detecting looming objects in dim light environments. The model inspired by the neural mechanism of insects' visual systems combines the DLMCE model with the classical four-layered LGMD1 model. It utilizes the outputs of the DLMCE model as inputs to feed the first layer of the LGMD1 model. Since the faint motion cues are enhanced by the DLMCE layer, the proposed model can recognize and extract the looming cues from very dark environments. Experimental results on the real dim-light dataset have demonstrated its effectiveness in looming motion detection. Moreover, the performance comparisons also show that the proposed model outperforms the comparative LGMD1 models in low light conditions.

In the future, we will explore the adaptive shift between the day and night mode of collision detection and the enhancement of looming sensitivity.

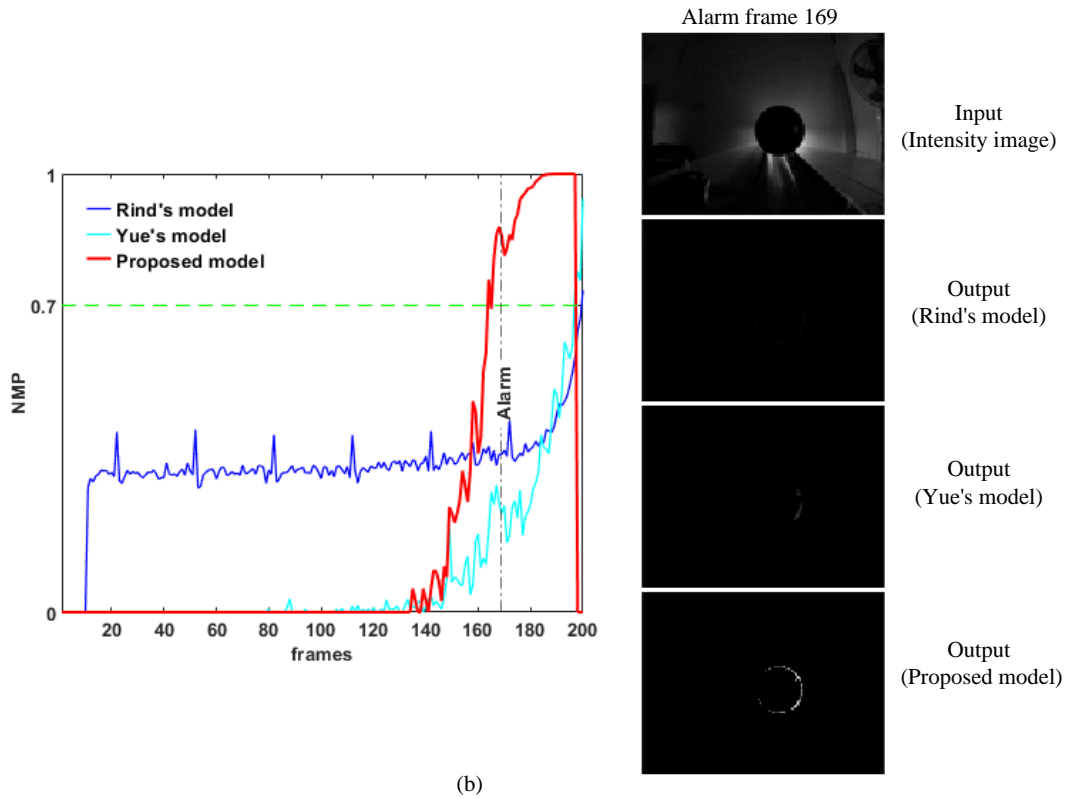
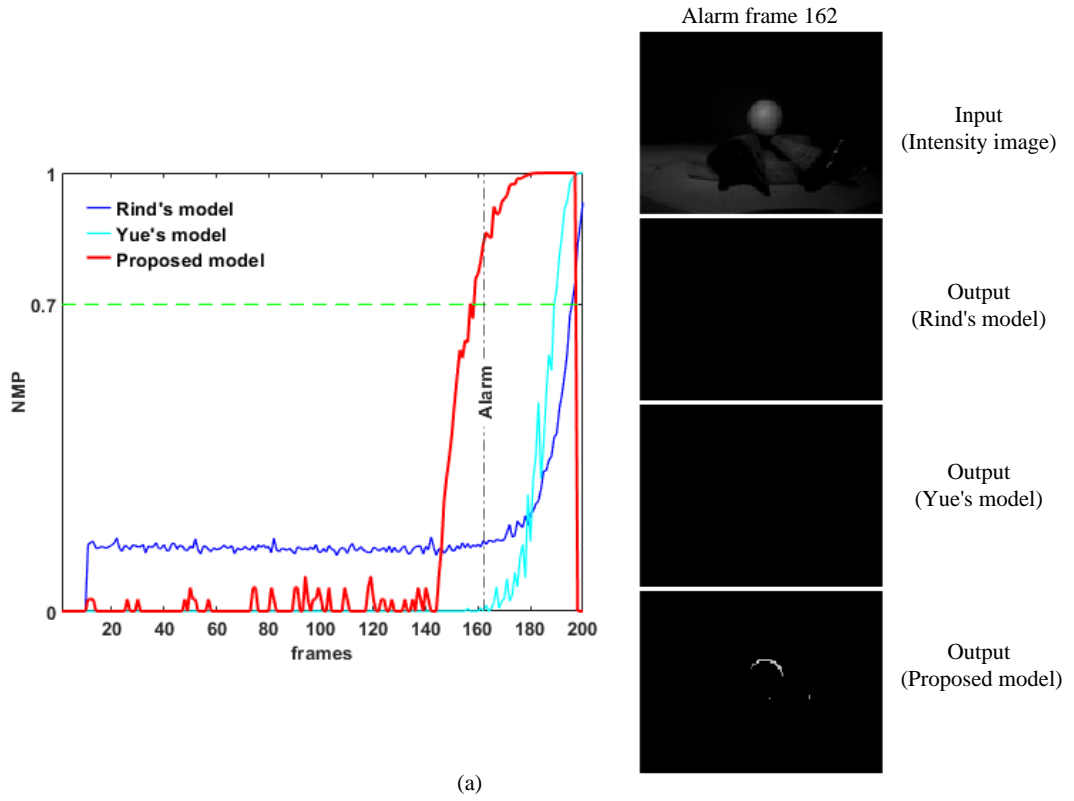


Figure 6.5: The comparison of LGMD1 neural response among Rind's model [5], Yue's model [7] and the proposed model for sequence one and two. The output image of the last layer in the model with respect to the alarm frame is also presented. Note that the alarm frame was produced by the proposed model. **The detected collision alarm frame in sequence one (a): Rind's model (Frame 199), Yue's model (Frame 192) and Proposed model (Frame 162). The detected collision alarm frame in sequence two (b): Rind's model (No alarm frame), Yue's model (Frame 200) and Proposed model (Frame 169).**

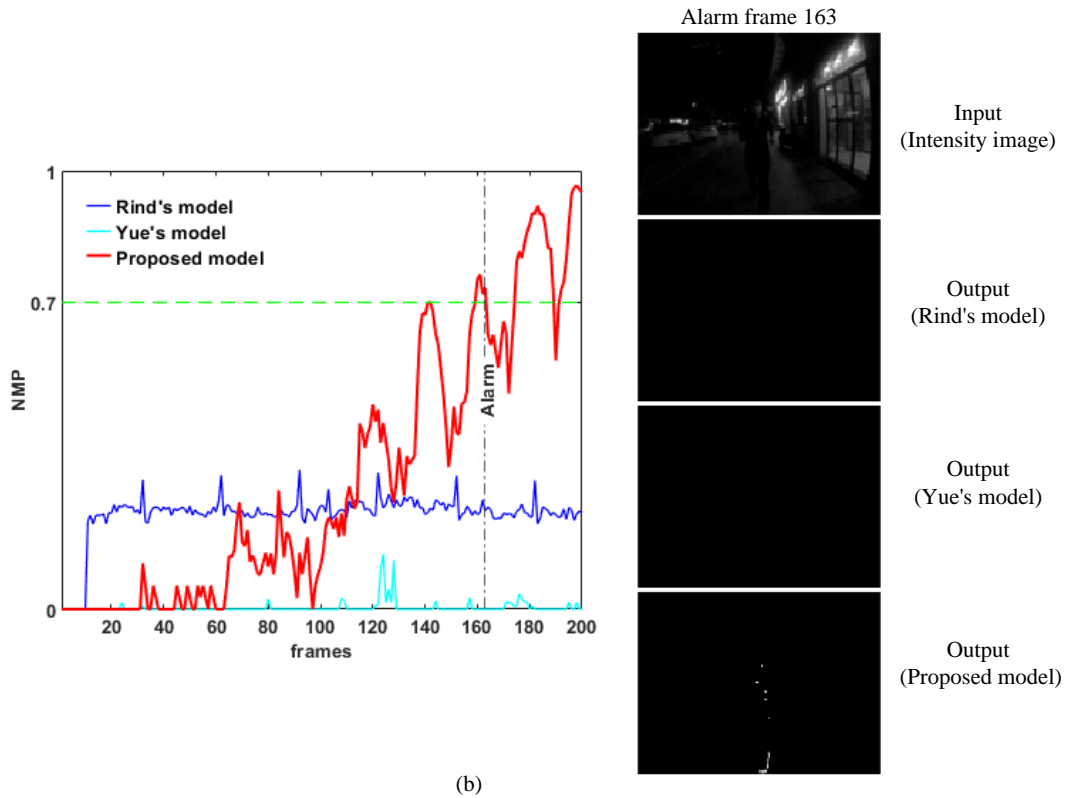
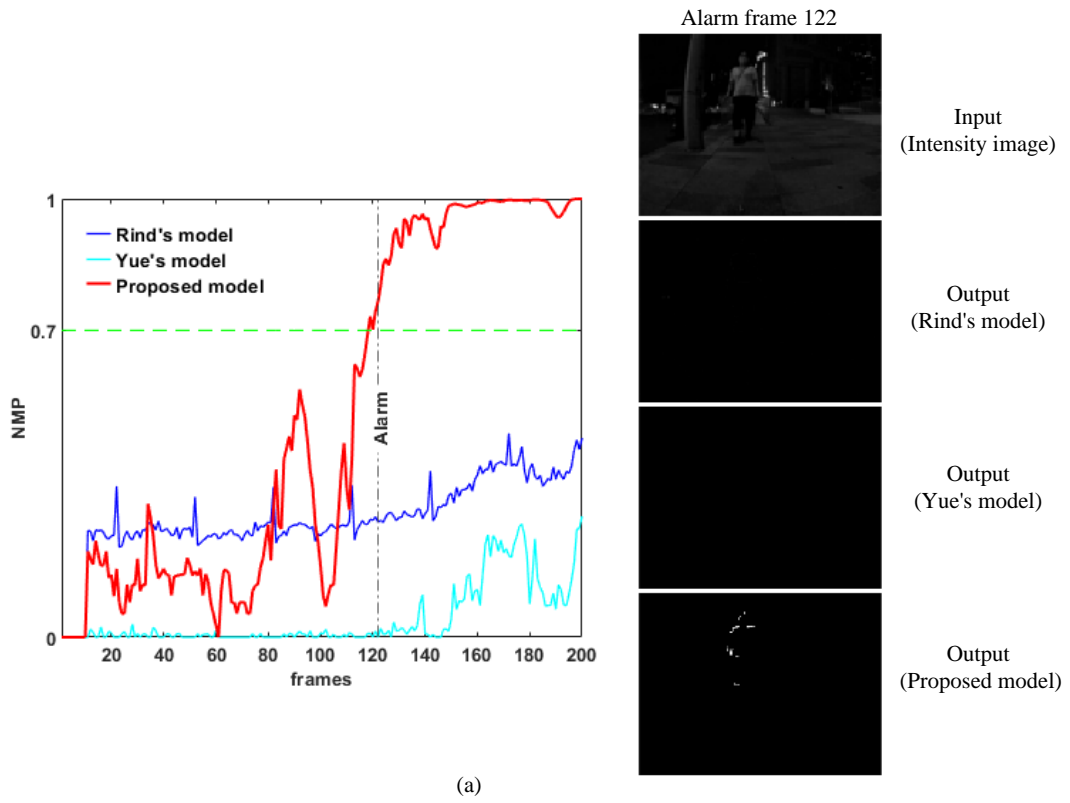


Figure 6.6: The comparison of LGMD1 neural response among Rind's model [5], Yue's model [7] and the proposed model for sequence three and four. The output image of the last layer in the model with respect to the alarm frame is also presented. Note that the alarm frame was produced by the proposed model. **The detected collision alarm frame in sequence three (a): Rind's model (No alarm frame), Yue's model (No alarm frame) and Proposed model (Frame 122). The detected collision alarm frame in sequence four (b): Rind's model (No alarm frame), Yue's model (No alarm frame) and Proposed model (Frame 163).**

Chapter 7

Research Contributions and Future Work

7.1 Research Contributions

This thesis mainly explores the LGMD1 based collision detection neural network model, the dark adaptation framework, the dim-light motion cues enhancement model, and the dim-light collision detection model. The proposed models primarily focus on mimicking the LGMD1's looming selectivity and the nocturnal animal's visual processing strategies in enhancing faint motion cues. We have provided rigorous mathematical descriptions and systematically tests for the proposed models. The main contributions of the thesis are summarized below.

- A new LGMD1 based neural network model is proposed to enhance the direction selectivity between looming and translating motion via neural competition. The neural competition based LGMD1 model within a framework of separated ON and OFF pathways can shut off the translating response, thus enhancing collision selectivity. Moreover, the denoising mechanism in the model guarantees reliable collision detection. Experimental results show that the proposed LGMD1 model is more robust in discriminating looming objects from translating ones compared with comparative models.
- A dark adaptation based framework is proposed to enhance low light images.

The proposed dark adaptation method applies a series of canonical neural computations to raise the intensities of dark regions whilst preserving the naturalness of illumination. For enhancing color image, the proposed framework combines the psychophysical power law with the photoreceptors' response properties to different wavelengths of light during the application of dark adaptation, including a comprehensive analysis of parameters for the R, G and B color channels. Experimental results have demonstrated that the proposed method is effective in preserving image naturalness while enhancing low light images.

- A dim-light motion cues enhancement (DLMCE) model is developed to enhance faint motion cues since the extraction of motion cues is essential for motion detection in low light scenes. The DLMCE model integrates dark-adaptation, spatio-temporal constraint and neural summation mechanisms achieved with canonical neural computations and visual neural summation. Experiments on synthetic and real datasets are presented to reveal the effectiveness of the proposed DLMCE model and show its superiority over the existing methods in terms of motion cues enhancement.
- A bio-inspired dim-light collision detection model is proposed for detecting looming objects in low light conditions. The model integrates the DLMCE model and the classical four-layered LGMD1 model where motion cues as inputs are enhanced to feed the first layer of the LGMD1. Experiments have demonstrated its effectiveness in detecting looming objects under low light conditions.

7.2 Future Work

In the natural world, animals' exquisite visual systems have provided us with a rich source of inspiration. It is therefore possible to develop robust artificial vision systems with specific neural processing strategies. The studies discussed in this thesis represent only a small step in modeling the collision selectivity of the LGMD1 neuron in the locust's brain and night vision mechanisms in nocturnal insects for applications.

The following research directions could be the future tasks for further exploration.

- Enhancing the LGMD1 model's looming selectivity between looming and receding. The effect of suppressing near receding objects in the proposed LGMD1 neural network model is limited. Exploring a more effective implementation way of the lateral inhibition mechanism may help enhance the collision selectivity when challenged with the near receding stimuli.
- Developing a new low-light image enhancement method by integrating neural denoising mechanisms. The proposed dark adaptation framework for enhancing low light images only involves intensity transformation processing. Therefore, combining with effective denoising strategies like the neural summation denoising may provide a solution to make results more visually pleasant.
- Modeling potential DLMCE based motion detection models. Since the DLMCE model enhances faint motion cues in low light conditions, it can be integrated into other directional selective neural network models to handle more complex motion recognition tasks besides integrated into the LGMD1 model.
- Exploring an adaptive shift between the day and night mode of collision detection. The proposed LGMD1 based collision detection models for normal illumination levels and dim-light environments are different as the manner that the neural information processed by the LGMD1 varies when the average intensity of the background has changed greatly. To improve the adaptation for collision detection, other neural information processing mechanisms could be simulated and integrated with the LGMD1 model to alter corresponding collision detection circuits according to the tremendous change in luminance levels.

Bibliography

- [1] F. Gabbiani, H. G. Krapp, N. Hatsopoulos, C.-H. Mo, C. Koch, and G. Laurent, “Multiplication and stimulus invariance in a looming-sensitive neuron,” *Journal of Physiology-Paris*, vol. 98, no. 1-3, pp. 19–34, 2004.
- [2] M. Frye, “Elementary motion detectors,” *Current Biology*, vol. 25, no. 6, pp. R215–R217, 2015.
- [3] A. Arenz, M. S. Drews, F. G. Richter, G. Ammer, and A. Borst, “The temporal tuning of the drosophila motion detectors is determined by the dynamics of their input elements,” *Current Biology*, vol. 27, no. 7, pp. 929–944, 2017.
- [4] F. C. Rind and P. J. Simmons, “Seeing what is coming: building collision-sensitive neurones,” *Trends in neurosciences*, vol. 22, no. 5, pp. 215–220, 1999.
- [5] F. C. Rind and D. Bramwell, “Neural network based on the input organization of an identified neuron signaling impending collision,” vol. 75, no. 3, pp. 967–985, 1996.
- [6] S. Yue and F. C. Rind, “Visual motion pattern extraction and fusion for collision detection in complex dynamic scenes,” vol. 104, no. 1, pp. 48–60, 2006.
- [7] S. Yue and F. C. Rind, “Collision detection in complex dynamic scenes using an lgmd-based visual neural network with feature enhancement,” vol. 17, no. 3, pp. 705–716, 2006.
- [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] F. Lei, Z. Peng, V. Cutsuridis, M. Liu, Y. Zhang, and S. Yue, “Competition between on and off neural pathways enhancing collision selectivity,” 2020.
- [10] E. J. Warrant, “The remarkable visual capacities of nocturnal insects: vision at the limits with small eyes and tiny brains,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1717, p. 20160063, 2017.
- [11] E. Warrant, M. Oskarsson, and H. Malm, “The remarkable visual abilities of nocturnal insects: neural principles and bioinspired night-vision algorithms,” *Proceedings of the IEEE*, vol. 102, no. 10, pp. 1411–1426, 2014.
- [12] T. Matic and S. Laughlin, “Changes in the intensity-response function of an insect’s photoreceptors due to light adaptation,” *Journal of comparative physiology*, vol. 145, no. 2, pp. 169–177, 1981.
- [13] R. A. Normann and I. Perlman, “The effects of background illumination on the photoresponses of red and green cones,” *The Journal of Physiology*, vol. 286, no. 1, pp. 491–507, 1979.
- [14] V. Bonin, V. Mante, and M. Carandini, “The suppressive field of neurons in lateral geniculate nucleus,” *Journal of Neuroscience*, vol. 25, no. 47, pp. 10844–10856, 2005.
- [15] M. Carandini and D. J. Heeger, “Normalization as a canonical neural computation,” *Nature Reviews Neuroscience*, vol. 13, no. 1, pp. 51–62, 2012.
- [16] B. Greiner, W. A. Ribi, and E. J. Warrant, “A neural network to improve dim-light vision? dendritic fields of first-order interneurons in the nocturnal bee *megalopecta genalis*,” *Cell and tissue research*, vol. 322, no. 2, pp. 313–320, 2005.
- [17] E. J. Warrant, “Seeing better at night: life style, eye design and the optimum strategy of spatial and temporal summation,” *Vision research*, vol. 39, no. 9, pp. 1611–1630, 1999.
- [18] W. Kim, “Moving object detection using edges of residuals under varying illuminations,” *Multi-media Systems*, vol. 25, no. 3, pp. 155–163, 2019.

- [19] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” *Graphics gems*, pp. 474–485, 1994.
- [20] X. Guo, Y. Li, and H. Ling, “Lime: Low-light image enhancement via illumination map estimation,” *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.
- [21] M. Tanaka, T. Shibata, and M. Okutomi, “Gradient-based low-light image enhancement,” in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–2, IEEE, 2019.
- [22] S. Yue and F. C. Rind, “A collision detection system for a mobile robot inspired by the locust visual system,” in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 3832–3837, IEEE, 2005.
- [23] K. Eichler, F. Li, A. Litwin-Kumar, Y. Park, I. Andrade, C. M. Schneider-Mizell, T. Saumweber, A. Huser, C. Eschbach, B. Gerber, *et al.*, “The complete connectome of a learning and memory centre in an insect brain,” *Nature*, vol. 548, no. 7666, pp. 175–182, 2017.
- [24] 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,” vol. 6, no. June, pp. 1–16, 2016.
- [25] R. B. Dewell and F. Gabbiani, “Biophysics of object segmentation in a collision-detecting neuron,” *ELife*, vol. 7, p. e34238, 2018.
- [26] 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 Transactions on Neural Networks and Learning Systems*, 2019.
- [27] A. L. Stöckl, D. C. O’Carroll, and E. J. Warrant, “Neural summation in the hawkmoth visual system extends the limits of vision in dim light,” *Current Biology*, vol. 26, no. 6, pp. 821–826, 2016.
- [28] E. Warrant and M. Dacke, “Visual navigation in nocturnal insects,” *Physiology*, vol. 31, no. 3, pp. 182–192, 2016.
- [29] P. J. Simmons and F. C. Rind, “Responses to object approach by a wide field visual neurone, the lgmd2 of the locust: Characterization and image cues,” vol. 180, no. 3, pp. 203–214, 1997.
- [30] M. Blanchard, F. C. Rind, and P. F. Verschure, “Collision avoidance using a model of the locust lgmd neuron,” *Robotics and Autonomous Systems*, vol. 30, no. 1-2, pp. 17–38, 2000.
- [31] C. Hu, F. Arvin, C. Xiong, and S. Yue, “Bio-inspired embedded vision system for autonomous micro-robots: the lgmd case,” *IEEE transactions on cognitive and developmental systems*, vol. 9, no. 3, pp. 241–254, 2016.
- [32] S. Yue, F. C. Rind, M. S. Keil, J. Cuadri, and R. Stafford, “A bio-inspired visual collision detection mechanism for cars: Optimisation of a model of a locust neuron to a novel environment,” *Neurocomputing*, vol. 69, no. 13-15, pp. 1591–1598, 2006.
- [33] F. Gabbiani and H. G. Krapp, “Spike-frequency adaptation and intrinsic properties of an identified, looming-sensitive neuron,” vol. 96, no. 6, pp. 2951–2962, 2006.
- [34] H. Ögmen and S. Gagné, “Neural models for sustained and on-off units of insect lamina,” vol. 63, no. 1, pp. 51–60, 1990.
- [35] D. Krotov and J. J. Hopfield, “Unsupervised learning by competing hidden units,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 16, pp. 7723–7731, 2019.
- [36] F. C. Rind and P. J. Simmons, “Orthopteran dcmd neuron: a reevaluation of responses to moving objects. i. selective responses to approaching objects,” *Journal of Neurophysiology*, vol. 68, no. 5, pp. 1654–1666, 1992.
- [37] M. S. Maisak *et al.*, “A directional tuning map of drosophila elementary motion detectors,” *Nature*, vol. 500, no. 7461, pp. 175–181, 2013.
- [38] S. Yue, R. D. Santer, Y. Yamawaki, and F. C. Rind, “Reactive direction control for a mobile robot: a locust-like control of escape direction emerges when a bilateral pair of model locust visual neurons are integrated,” *Autonomous Robots*, vol. 28, no. 2, p. 151, 2010.
- [39] H. Wang, J. Peng, and S. Yue, “A directionally selective small target motion detecting visual neural network in cluttered backgrounds,” pp. 1–14, 2018.

- [40] 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.
- [41] 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," 2019.
- [42] H. Malm, M. Oskarsson, E. Warrant, P. Clarberg, J. Hasselgren, and C. Lejdfors, "Adaptive enhancement and noise reduction in very low light-level video," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.
- [43] E. Warrant, M. Oskarsson, and H. Malm, "A night vision algorithm inspired by the visual system of a nocturnal bee," in *Biomimetics in Photonics*, pp. 178–203, CRC Press, 2012.
- [44] M. O'Shea and J. Williams, "The anatomy and output connection of a locust visual interneurone; the lobular giant movement detector (lgmd) neurone," *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 91, no. 3, pp. 257–266, 1974.
- [45] F. Gabbiani, H. G. Krapp, C. Koch, and G. Laurent, "Multiplicative computation in a visual neuron sensitive to looming," *Nature*, vol. 420, no. 6913, pp. 320–324, 2002.
- [46] P. D. Barnett, K. Nordström, and D. C. O'carroll, "Retinotopic organization of small-field-target-detecting neurons in the insect visual system," *Current Biology*, vol. 17, no. 7, pp. 569–578, 2007.
- [47] A. Borst and J. Haag, "Neural networks in the cockpit of the fly," *Journal of Comparative Physiology A*, vol. 188, no. 6, pp. 419–437, 2002.
- [48] F. Iida and D. Lambrinos, "Navigation in an autonomous flying robot by using a biologically inspired visual odometer," in *Sensor fusion and decentralized control in robotic systems III*, vol. 4196, pp. 86–97, International Society for Optics and Photonics, 2000.
- [49] A. S. Mauss, M. Meier, E. Serbe, and A. Borst, "Optogenetic and pharmacologic dissection of feedforward inhibition in drosophila motion vision," vol. 34, no. 6, pp. 2254–2263, 2014.
- [50] 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, p. 300, 2010.
- [51] S. Yue and Q. Fu, "Modeling direction selective visual neural network with on and off pathways for extracting motion cues from cluttered background," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 831–838, IEEE, 2017.
- [52] N. Franceschini, A. Riehle, and A. Le Nestour, "Directionally selective motion detection by insect neurons," in *Facets of vision*, pp. 360–390, Springer, 1989.
- [53] A. Borst and M. Egelhaaf, "Detecting visual motion: Theory and models, visual motion and its role in the stabilization of gaze. fa miles and j. wallman," 1993.
- [54] R. R. Harrison, "A biologically inspired analog ic for visual collision detection," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 11, pp. 2308–2318, 2005.
- [55] H. Okuno and T. Yagi, "Bio-inspired real-time robot vision for collision avoidance," *Journal of Robotics and Mechatronics*, vol. 20, no. 1, p. 68, 2008.
- [56] M. Sarkar, D. S. S. Bello, C. van Hoof, and A. Theuvsen, "A biologically inspired collision detection algorithm using differential optic flow imaging," in *2010 Biomedical Circuits and Systems Conference (BioCAS)*, pp. 250–253, IEEE, 2010.
- [57] R. Stafford, R. D. Santer, and F. C. Rind, "A bio-inspired visual collision detection mechanism for cars: combining insect inspired neurons to create a robust system," *BioSystems*, vol. 87, no. 2-3, pp. 164–171, 2007.
- [58] M. B. Milde, O. J. Bertrand, R. Benosman, M. Egelhaaf, and E. Chicca, "Bioinspired event-driven collision avoidance algorithm based on optic flow," in *2015 International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, pp. 1–7, IEEE, 2015.
- [59] 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.
- [60] Y. Wang and B. J. Frost, "Time to collision is signalled by neurons in the nucleus rotundus of pigeons," *Nature*, vol. 356, no. 6366, pp. 236–238, 1992.

- [61] T. A. Münch, R. A. Da Silveira, S. Siegert, T. J. Viney, G. B. Awatramani, and B. Roska, "Approach sensitivity in the retina processed by a multifunctional neural circuit," *Nature neuroscience*, vol. 12, no. 10, pp. 1308–1316, 2009.
- [62] F. Gabbiani, H. G. Krapp, and G. Laurent, "Computation of object approach by a wide-field, motion-sensitive neuron," vol. 19, no. 3, pp. 1122–1141, 1999.
- [63] H. Sun and B. J. Frost, "Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons," *Nature neuroscience*, vol. 1, no. 4, pp. 296–303, 1998.
- [64] T. Fülöp and Á. Zarándy, "Bio-inspired looming direction detection method," in *2012 13th International Workshop on Cellular Nanoscale Networks and their Applications*, pp. 1–6, IEEE, 2012.
- [65] Á. Zarándy and T. Fülöp, "Approaching object detector mouse retina circuit model analysis and implementation on cellular sensor-processor array," *International Journal of Circuit Theory and Applications*, vol. 40, no. 12, pp. 1249–1264, 2012.
- [66] P. J. Simmons and F. C. Rind, "Orthopteran dcmd neuron: a reevaluation of responses to moving objects. ii. critical cues for detecting approaching objects," *Journal of neurophysiology*, vol. 68, no. 5, pp. 1667–1682, 1992.
- [67] Y. Zhu, R. B. Dewell, H. Wang, and F. Gabbiani, "Pre-synaptic muscarinic excitation enhances the discrimination of looming stimuli in a collision-detection neuron," *Cell reports*, vol. 23, no. 8, pp. 2365–2378, 2018.
- [68] B. B. Guest and J. R. Gray, "Responses of a looming-sensitive neuron to compound and paired object approaches," *Journal of neurophysiology*, vol. 95, no. 3, pp. 1428–1441, 2006.
- [69] J. M. Yakubowski, G. A. McMillan, and J. R. Gray, "Background visual motion affects responses of an insect motion-sensitive neuron to objects deviating from a collision course," *Physiological Reports*, vol. 4, no. 10, p. e12801, 2016.
- [70] F. C. Rind, "A chemical synapse between two motion detecting neurones in the locust brain," *Journal of Experimental Biology*, vol. 110, no. 1, pp. 143–167, 1984.
- [71] F. C. Rind and G. Leitterer, "Immunocytochemical evidence that collision sensing neurons in the locust visual system contain acetylcholine," *Journal of Comparative Neurology*, vol. 423, no. 3, pp. 389–401, 2000.
- [72] S. Yue and F. C. Rind, "Near range path navigation using lgmd visual neural networks," in *2009 2nd IEEE International Conference on Computer Science and Information Technology*, pp. 105–109, IEEE, 2009.
- [73] F. Gabbiani, G. Laurent, N. Hatsopoulos, and H. G. Krapp, "The many ways of building collision-sensitive neurons," *Trends in neurosciences*, vol. 22, no. 10, pp. 437–438, 1999.
- [74] S. B. i Badia, U. Bernardet, and P. F. Verschure, "Non-linear neuronal responses as an emergent property of afferent networks: A case study of the locust lobula giant movement detector," vol. 6, no. 3, pp. 1–15, 2010.
- [75] M. Keil, "Emergence of multiplication in a biophysical model of a wide-field visual neuron for computing object approaches: Dynamics, peaks, & fits," *Advances in Neural Information Processing Systems*, vol. 24, pp. 469–477, 2011.
- [76] M. S. Keil, "Dendritic pooling of noisy threshold processes can explain many properties of a collision-sensitive visual neuron," *PLoS computational biology*, vol. 11, no. 10, p. e1004479, 2015.
- [77] S. P. Peron and F. Gabbiani, "Role of spike-frequency adaptation in shaping neuronal response to dynamic stimuli," vol. 100, no. 6, pp. 505–520, 2009.
- [78] S. Peron and F. Gabbiani, "Spike frequency adaptation mediates looming stimulus selectivity in a collision-detecting neuron," vol. 12, no. 3, p. 318, 2009.
- [79] H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 208, no. 1173, pp. 385–397, 1980.

- [80] J. M. Galbraith, G. T. Kenyon, and R. W. Ziolkowski, "Time-to-collision estimation from motion based on primate visual processing," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1279–1291, 2005.
- [81] J. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images-a review," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 917–935, 1988.
- [82] J. J. Gibson, "On the analysis of change in the optic array.," *Scandinavian Journal of Psychology*, 1977.
- [83] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [84] D. Fleet and Y. Weiss, "Optical flow estimation," in *Handbook of mathematical models in computer vision*, pp. 237–257, Springer, 2006.
- [85] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International journal of computer vision*, vol. 12, no. 1, pp. 43–77, 1994.
- [86] B. D. Lucas, T. Kanade, *et al.*, "An iterative image registration technique with an application to stereo vision," Vancouver, British Columbia, 1981.
- [87] B. D. Lucas, *Generalized image matching by the method of differences*. Carnegie Mellon University, 1985.
- [88] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: Combining local and global optic flow methods," *International journal of computer vision*, vol. 61, no. 3, pp. 211–231, 2005.
- [89] J.-Y. Bouguet *et al.*, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel corporation*, vol. 5, no. 1-10, p. 4, 2001.
- [90] K. Souhila and A. Karim, "Optical flow based robot obstacle avoidance," *International Journal of Advanced Robotic Systems*, vol. 4, no. 1, p. 2, 2007.
- [91] H. W. Ho, G. C. de Croon, and Q. Chu, "Distance and velocity estimation using optical flow from a monocular camera," *International Journal of Micro Air Vehicles*, vol. 9, no. 3, pp. 198–208, 2017.
- [92] D. Coombs, M. Herman, T.-H. Hong, and M. Nashman, "Real-time obstacle avoidance using central flow divergence, and peripheral flow," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 1, pp. 49–59, 1998.
- [93] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1508–1515, Ieee, 2005.
- [94] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European conference on computer vision*, pp. 430–443, Springer, 2006.
- [95] T. Mori and S. Scherer, "First results in detecting and avoiding frontal obstacles from a monocular camera for micro unmanned aerial vehicles," in *2013 IEEE International Conference on Robotics and Automation*, pp. 1750–1757, IEEE, 2013.
- [96] A. Al-Kaff, F. García, D. Martín, A. De La Escalera, and J. Armingol, "Obstacle detection and avoidance system based on monocular camera and size expansion algorithm for uavs," *Sensors*, vol. 17, no. 5, p. 1061, 2017.
- [97] A. Chavez and D. Gustafson, "Vision-based obstacle avoidance using sift features," in *International Symposium on Visual Computing*, pp. 550–557, Springer, 2009.
- [98] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, pp. 1150–1157, Ieee, 1999.
- [99] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [100] N. Y. Khan, B. McCane, and G. Wyvill, "Sift and surf performance evaluation against various image deformations on benchmark dataset," in *2011 International Conference on Digital Image Computing: Techniques and Applications*, pp. 501–506, IEEE, 2011.

- [101] E. Warrant and M. Dacke, "Vision and visual navigation in nocturnal insects," *Annual review of entomology*, vol. 56, pp. 239–254, 2011.
- [102] A. A. Lazar, N. H. Ukani, and Y. Zhou, "Sparse identification of contrast gain control in the fruit fly photoreceptor and amacrine cell layer," *The Journal of Mathematical Neuroscience*, vol. 10, no. 1, pp. 1–35, 2020.
- [103] P. Ala-Laurila, "Visual neuroscience: How do moths see to fly at night?," *Current Biology*, vol. 26, no. 6, pp. R231–R233, 2016.
- [104] E. Warrant, "Vision in the dimmest habitats on earth," *Journal of Comparative Physiology A*, vol. 190, no. 10, pp. 765–789, 2004.
- [105] B. Greiner, "Adaptations for nocturnal vision in insect apposition eyes," *International review of cytology*, vol. 250, pp. 1–46, 2006.
- [106] A. Stöckl, *Neurons against noise: neural adaptations for dim light vision in hawkmoths*. PhD thesis, Lund University, 2016.
- [107] L. E. Lipetz, "The relation of physiological and psychological aspects of sensory intensity," in *Principles of receptor physiology*, pp. 191–225, Springer, 1971.
- [108] D. MacKay, "Psychophysics of perceived intensity: A theoretical basis for fechner's and stevens' laws," *Science*, vol. 139, no. 3560, pp. 1213a–1216, 1963.
- [109] G. Sperling, "Model of visual adaptation and contrast detection," *Perception & Psychophysics*, vol. 8, no. 3, pp. 143–157, 1970.
- [110] A. Honkanen, E.-V. Immonen, I. Salmela, K. Heimonen, and M. Weckström, "Insect photoreceptor adaptations to night vision," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1717, p. 20160077, 2017.
- [111] T. Brosch and H. Neumann, "Computing with a canonical neural circuits model with pool normalization and modulating feedback," *Neural computation*, vol. 26, no. 12, pp. 2735–2789, 2014.
- [112] L. Busse, A. R. Wade, and M. Carandini, "Representation of concurrent stimuli by population activity in visual cortex," *Neuron*, vol. 64, no. 6, pp. 931–942, 2009.
- [113] J. Y. Angela, M. A. Giese, and T. A. Poggio, "Biophysically plausible implementations of the maximum operation," *Neural computation*, vol. 14, no. 12, pp. 2857–2881, 2002.
- [114] J. H. Reynolds and D. J. Heeger, "The normalization model of attention," *Neuron*, vol. 61, 2009.
- [115] S. R. Olsen, V. Bhandawat, and R. I. Wilson, "Divisive normalization in olfactory population codes," *Neuron*, vol. 66, no. 2, pp. 287–299, 2010.
- [116] E. J. Warrant, "Superior visual performance in nocturnal insects: neural principles and bio-inspired technologies," in *Bioinspiration, Biomimetics, and Bioreplication 2016*, vol. 9797, p. 979703, International Society for Optics and Photonics, 2016.
- [117] A. Klaus and E. J. Warrant, "Optimum spatiotemporal receptive fields for vision in dim light," *Journal of vision*, vol. 9, no. 4, pp. 18–18, 2009.
- [118] J. C. Theobald, B. Greiner, W. T. Wcislo, and E. J. Warrant, "Visual summation in night-flying sweat bees: a theoretical study," *Vision research*, vol. 46, no. 14, pp. 2298–2309, 2006.
- [119] H. Malm and E. Warrant, "Motion dependent spatiotemporal smoothing for noise reduction in very dim light image sequences," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 954–959, IEEE, 2006.
- [120] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [121] P. E. Trahanias and A. N. Venetsanopoulos, "Color image enhancement through 3-d histogram equalization," in *11th IAPR International Conference on Pattern Recognition. Vol. III. Conference C: Image, Speech and Signal Analysis*, vol. 1, pp. 545–548, IEEE Computer Society, 1992.
- [122] H. Lin and Z. Shi, "Multi-scale retinex improvement for nighttime image enhancement," *Optik*, vol. 125, no. 24, pp. 7143–7148, 2014.

- [123] T. Shibata, M. Tanaka, and M. Okutomi, "Gradient-domain image reconstruction framework with intensity-range and base-structure constraints," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2745–2753, 2016.
- [124] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.
- [125] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3291–3300, 2018.
- [126] F. Lv, F. Lu, J. Wu, and C. Lim, "Mbllen: Low-light image/video enhancement using cnns.," in *BMVC*, vol. 220, p. 4, 2018.
- [127] H. Jiang and Y. Zheng, "Learning to see moving objects in the dark," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7324–7333, 2019.
- [128] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, 2007.
- [129] G. Zahi and S. Yue, "Automatic detection of low light images in a video sequence shot under different light conditions," in *2013 European Modelling Symposium*, pp. 271–276, IEEE, 2013.
- [130] R. C. Gonzalez, R. E. Woods, *et al.*, "Digital image processing," 2002.
- [131] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.
- [132] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE transactions on image processing*, vol. 6, no. 3, pp. 451–462, 1997.
- [133] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image processing*, vol. 6, no. 7, pp. 965–976, 1997.
- [134] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.
- [135] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.
- [136] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2782–2790, 2016.
- [137] J. M. DiCarlo and B. A. Wandell, "Rendering high dynamic range images," in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, vol. 3965, pp. 392–401, International Society for Optics and Photonics, 2000.
- [138] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pp. 249–256, 2002.
- [139] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "Llcn: A convolutional neural network for low-light image enhancement," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, IEEE, 2017.
- [140] J. S. Humbert and A. M. Hyslop, "Bioinspired visuomotor convergence," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 121–130, 2009.
- [141] S. Sponberg, J. P. Dyhr, R. W. Hall, and T. L. Daniel, "Luminance-dependent visual processing enables moth flight in low light," *Science*, vol. 348, no. 6240, pp. 1245–1248, 2015.
- [142] R. Behnia, D. A. Clark, A. G. Carter, T. R. Clandinin, and C. Desplan, "Processing properties of on and off pathways for drosophila motion detection," *Nature*, vol. 512, no. 7515, pp. 427–430, 2014.
- [143] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE transactions on image processing*, vol. 14, no. 3, pp. 294–307, 2005.

- [144] C. Zhan, X. Duan, S. Xu, Z. Song, and M. Luo, "An improved moving object detection algorithm based on frame difference and edge detection," in *Fourth international conference on image and graphics (ICIG 2007)*, pp. 519–523, IEEE, 2007.
- [145] I. Saleemi and M. Shah, "Multiframe many-many point correspondence for vehicle tracking in high density wide area aerial videos," *International journal of computer vision*, vol. 104, no. 2, pp. 198–219, 2013.
- [146] P. Sundberg, T. Brox, M. Maire, P. Arbeláez, and J. Malik, "Occlusion boundary detection and figure/ground assignment from optical flow," in *CVPR 2011*, pp. 2233–2240, IEEE, 2011.
- [147] A. Papazoglou and V. Ferrari, "Fast object segmentation in unconstrained video," in *Proceedings of the IEEE international conference on computer vision*, pp. 1777–1784, 2013.
- [148] Y.-H. Tsai, M.-H. Yang, and M. J. Black, "Video segmentation via object flow," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3899–3908, 2016.
- [149] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *European conference on computer vision*, pp. 282–295, Springer, 2010.
- [150] N. Sundaram, T. Brox, and K. Keutzer, "Dense point trajectories by gpu-accelerated large displacement optical flow," in *European conference on computer vision*, pp. 438–451, Springer, 2010.
- [151] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 500–513, 2010.
- [152] F. C. Rind, "A directionally selective motion-detecting neurone in the brain of the locust: physiological and morphological characterization," *Journal of Experimental Biology*, vol. 149, no. 1, pp. 1–19, 1990.
- [153] S. Yue and F. C. Rind, "Postsynaptic organizations of directional selective visual neural networks for collision detection," *Neurocomputing*, vol. 103, pp. 50–62, 2013.
- [154] 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, 2016.
- [155] S. Yue and F. C. Rind, "A synthetic vision system using directionally selective motion detectors to recognize collision," *Artificial life*, vol. 13, no. 2, pp. 93–122, 2007.
- [156] F. Lei, Z. Peng, M. Liu, J. Peng, V. Cutsuridis, and S. Yue, "A robust visual system for looming cue detection against translating motion," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [157] H. Wang, Q. Fu, H. Wang, J. Peng, P. Baxter, C. Hu, and S. Yue, "Angular velocity estimation of image motion mimicking the honeybee tunnel centring behaviour," 2019.
- [158] F. C. Rind and P. J. Simmons, "Local circuit for the computation of object approach by an identified visual neuron in the locust," vol. 395, no. 3, pp. 405–415, 1998.
- [159] P. W. Jones and F. Gabbiani, "Synchronized neural input shapes stimulus selectivity in a collision-detecting neuron," *Current Biology*, vol. 20, no. 22, pp. 2052–2057, 2010.
- [160] F. C. Rind, R. D. Santer, J. M. Blanchard, and P. F. Verschure, "Locust's looming detectors for robot sensors," in *Sensors and sensing in biology and engineering*, pp. 237–250, Springer, 2003.
- [161] W. Catton, "The effects of stimulus area and intensity on the on/off ratio of some locust visual interneurons," *Journal of Insect Physiology*, vol. 28, no. 3, pp. 285–291, 1982.
- [162] D. Osorio, "Mechanisms of early visual processing in the medulla of the locust optic lobe: how self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells," vol. 7, no. 4, pp. 345–355, 1991.
- [163] R. D. Santer, R. Stafford, and F. C. Rind, "Retinally-generated saccadic suppression of a locust looming-detector neuron: investigations using a robot locust," vol. 1, no. 1, pp. 61–77, 2004.
- [164] A. D. Straw, "Vision egg: an open-source library for realtime visual stimulus generation," *Frontiers in neuroinformatics*, vol. 2, p. 4, 2008.

- [165] E. D. Micheli, V. Torre, and S. Uras, “The accuracy of the computation of optical flow and of the recovery of motion parameters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 5, pp. 434–447, 1993.
- [166] M. Zaretsky and C. F. ROWELL, “Saccadic suppression by corollary discharge in the locust,” *Nature*, vol. 280, no. 5723, pp. 583–585, 1979.
- [167] H. H. Thoen, J. Marshall, G. H. Wolff, and N. J. Strausfeld, “Insect-like organization of the stomatopod central complex: functional and phylogenetic implications,” *Frontiers in Behavioral Neuroscience*, vol. 11, p. 12, 2017.
- [168] M. Aladem, S. Baek, and S. A. Rawashdeh, “Evaluation of image enhancement techniques for vision-based navigation under low illumination,” *Journal of Robotics*, vol. 2019, 2019.
- [169] R. V. Frolov and I. I. Ignatova, “Electrophysiological adaptations of insect photoreceptors and their elementary responses to diurnal and nocturnal lifestyles,” *Journal of Comparative Physiology A*, vol. 206, no. 1, pp. 55–69, 2020.
- [170] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, “Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms,” *Journal of Digital imaging*, vol. 11, no. 4, p. 193, 1998.
- [171] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based hdr reconstruction of dynamic scenes,” *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203–1, 2012.
- [172] Z. Ying, G. Li, and W. Gao, “A bio-inspired multi-exposure fusion framework for low-light image enhancement,” *arXiv preprint arXiv:1711.00591*, 2017.
- [173] J. C. Stevens and L. E. Marks, “Stevens power law in vision: exponents, intercepts, and thresholds,” *Fechner Day*, vol. 99, pp. 82–87, 1999.
- [174] S. J. Mitchell and R. A. Silver, “Shunting inhibition modulates neuronal gain during synaptic excitation,” *Neuron*, vol. 38, no. 3, pp. 433–445, 2003.
- [175] B. Wark, B. N. Lundstrom, and A. Fairhall, “Sensory adaptation,” *Current opinion in neurobiology*, vol. 17, no. 4, pp. 423–429, 2007.
- [176] N. G. Paulter, D. R. Larson, and J. J. Blair, “The iee standard on transitions, pulses, and related waveforms, std-181-2003,” *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 4, pp. 1209–1217, 2004.
- [177] T. Ohshiro, D. E. Angelaki, and G. C. DeAngelis, “A normalization model of multisensory integration,” *Nature neuroscience*, vol. 14, no. 6, pp. 775–782, 2011.
- [178] O. Schwartz and E. P. Simoncelli, “Natural sound statistics and divisive normalization in the auditory system,” *Advances in neural information processing systems*, pp. 166–172, 2001.
- [179] M. Rivera-Alba, S. N. Vitaladevuni, Y. Mishchenko, Z. Lu, S.-y. Takemura, L. Scheffer, I. A. Meinertzhagen, D. B. Chklovskii, and G. G. de Polavieja, “Wiring economy and volume exclusion determine neuronal placement in the drosophila brain,” *Current Biology*, vol. 21, no. 23, pp. 2000–2005, 2011.
- [180] P. Skorupski and L. Chittka, “Photoreceptor spectral sensitivity in the bumblebee, *bombus impatiens* (hymenoptera: Apidae),” *PLoS One*, vol. 5, no. 8, p. e12049, 2010.
- [181] N. Brenner, W. Bialek, and R. d. R. Van Steveninck, “Adaptive rescaling maximizes information transmission,” *Neuron*, vol. 26, no. 3, pp. 695–702, 2000.
- [182] S. Chen and A. Beghdadi, “Natural rendering of color image based on retinex,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 1813–1816, IEEE, 2009.
- [183] A. Borst and T. Euler, “Seeing things in motion: models, circuits, and mechanisms,” *Neuron*, vol. 71, no. 6, pp. 974–994, 2011.
- [184] J. Zhao, H. Wang, N. Bellotto, C. Hu, J. Peng, and S. Yue, “Enhancing lgmd’s looming selectivity for uav with spatial-temporal distributed presynaptic connections,” *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

- [185] Q. Fu, C. Hu, T. Liu, and S. Yue, "Collision selective lgmds neuron models research benefits from a vision-based autonomous micro robot," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3996–4002, IEEE, 2017.
- [186] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 5, pp. 694–711, 2006.
- [187] A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, pp. 208–212, 2011.
- [188] A. Danielyan, V. Katkovnik, and K. Egiazarian, "Bm3d frames and variational image deblurring," *IEEE Transactions on image processing*, vol. 21, no. 4, pp. 1715–1728, 2011.
- [189] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms," *IEEE Transactions on image processing*, vol. 21, no. 9, pp. 3952–3966, 2012.
- [190] A. I. Weber, K. Krishnamurthy, and A. L. Fairhall, "Coding principles in adaptation," *Annual review of vision science*, vol. 5, pp. 427–449, 2019.
- [191] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "From contours to regions: An empirical evaluation," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2294–2301, IEEE, 2009.
- [192] H. Wang, H. Wang, J. Zhao, C. Hu, J. Peng, and S. Yue, "A time-delay feedback neural network for discriminating small, fast-moving targets in complex dynamic environments," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [193] G. Schlotterer, "Response of the locust descending movement detector neuron to rapidly approaching and withdrawing visual stimuli," *Canadian Journal of Zoology*, vol. 55, no. 8, pp. 1372–1376, 1977.
- [194] F. C. Rind, "Non-directional, movement sensitive neurones of the locust optic lobe," vol. 161, no. 3, pp. 477–494, 1987.
- [195] D. C. O'Carroll and E. J. Warrant, "Vision in dim light: highlights and challenges," 2017.
- [196] A. L. Stöckl, D. C. O'Carroll, and E. J. Warrant, "Hawkmoth lamina monopolar cells act as dynamic spatial filters to optimize vision at different light levels," *Science Advances*, vol. 6, no. 16, p. eaaz8645, 2020.
- [197] H. Meng, S. Yue, A. Hunter, K. Appiah, M. Hobden, N. Priestley, P. Hobden, and C. Pettit, "A modified neural network model for lobula giant movement detector with additional depth movement feature," in *2009 International Joint Conference on Neural Networks*, pp. 2078–2083, IEEE, 2009.