Journal Pre-proof

A bio-inspired visual collision detection network integrated with dynamic temporal variance feedback regulated by scalable functional countering jitter streaming

Zefang Chang, Hao Chen, Mu Hua, Qinbing Fu, Jigen Peng

PII: S0893-6080(24)00811-6
DOI: https://doi.org/10.1016/j.neunet.2024.106882
Reference: NN 106882

To appear in: *Neural Networks*

Received date : 14 May 2024
Revised date : 29 October 2024
Accepted date : 31 October 2024

Please cite this article as: Z. Chang, H. Chen, M. Hua et al., A bio-inspired visual collision detection network integrated with dynamic temporal variance feedback regulated by scalable functional countering jitter streaming. *Neural Networks* (2024), doi: https://doi.org/10.1016/j.neunet.2024.106882.

A Bio-inspired Visual Collision Detection Network Integrated With Dynamic Temporal Variance Feedback Regulated by Scalable Functional Countering Jitter Streaming

Zefang Chang[a,b,1,*], Hao Chen[b,1], Mu Hua[c], Qinbing Fu[b], Jigen Peng[b]

[a]*Institute for Math & AI, Wuhan University,*
[b]*Machine Life and Intelligence Research Centre, School of Mathematics and Information Science, Guangzhou University,*
[c]*School of Computing and Mathematical Sciences, College of Science and Engineering, University of Leicester.*

Abstract

In pursuing artificial intelligence for efficient collision avoidance in robots, researchers draw inspiration from the locust's visual looming-sensitive neural circuit to establish an efficient neural network for collision detection. However, existing bio-inspired collision detection neural networks encounter challenges posed by jitter streaming, a phenomenon commonly experienced, for example, when a ground robot moves across uneven terrain. Visual inputs from jitter streaming induce significant fluctuations in grey values, distracting existing bio-inspired networks from extracting visually looming features. To overcome this limitation, we derive inspiration from the potential of feedback loops to enable the brain to generate a coherent visual perception. We introduce a novel dynamic temporal variance feedback loop regulated by scalable functional into the traditional bio-inspired collision detection neural network. This feedback mechanism extracts dynamic temporal variance information from the output of higher-order neurons in the conventional network to assess the fluctuation level of local neural responses and regulate it by a scalable functional to differentiate variance induced by incoherent visual input. Then the regulated signal is reintegrated into the input through negative feedback loop to reduce the incoherence of the signal within the network. Numerical experiments substantiate the effectiveness of the proposed feedback loop in promoting collision detection against jitter streaming. This study extends the capabilities of bio-inspired collision detection neural networks to address jitter streaming challenges, offering a novel insight into the potential of feedback mechanisms in enhancing visual neural abilities.

Keywords: Bio-inspired, Jitter streaming, LGMD, Dynamic temporal variance, Feedback

1. Introduction

Intelligent robots play a crucial role in enhancing production efficiency, quality and safety while simultaneously providing increased convenience and innovation, thus fostering societal development and progress. Visual collision perception technology is pivotal in enabling robots to identify and predict potential collision risks, thereby preventing accidents. This enhances the safety, reliability, and interactive capabilities of robots with their environment and humans and contributes to overall system efficiency. Building an artificial visual system for collision perception presents a significant challenge, particularly in achieving rapid and computationally-efficient collision detection against dynamic and complex backgrounds. Fortunately, despite its limited brain resources, the locust possesses remarkable visual collision avoidance capabilities, serving as a natural template for constructing intelligent units designed for robots' collision-free tasks[1, 2, 3].

Biological research on the locust visual system has unveiled a set of lobula giant movement detectors (LGMDs)[4, 5, 6, 7, 8], endowing them with the ability to process visual information swiftly, exhibit heightened sensitivity to frontal collision threats, and precisely capture and recognize details. Drawing from Oshea and Williams' theoretical framework for a more profound comprehension of the mechanisms of visual perception and the attributes of visual information in locust neural pathways[4], Rind and Bramwell[9] proposed that detecting brightness changes in the edges of moving target is pivotal for LGMD neurons to recognize radial motion patterns, as delineated in their exploration of LGMD neuron characteristics. Expanding on this information processing mechanism, Rind and Yue introduced the innovative classic LGMD model[10] for collision detection. In 2013, they[11] further refined the functional properties of this model and applied the enhanced neural network to robots.

Indeed, some pioneer work have been made to carry out LGMD model on a mobile robot platform and achieved good performances in specific scenarios [12, 13, 14]. However, the current implementation of the LGMD model on

*Corresponding author
Email address: zfchang@whu.edu.cn (Zefang Chang)
[1]These authors contributed equally to this work.

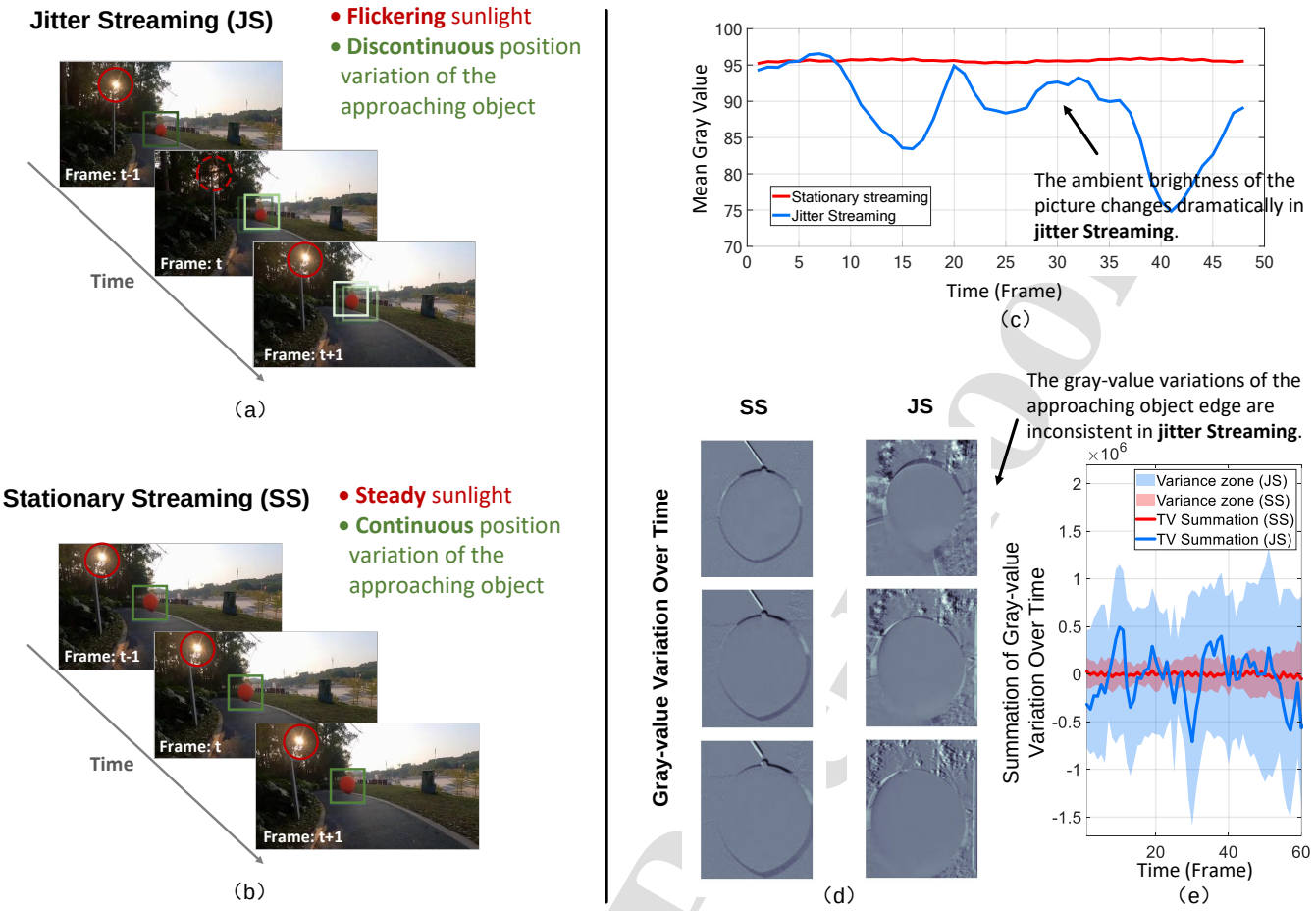

Figure 1: Illustrations of comparisons between jitter streaming and stationary streaming and issues induced by jitter streaming. (a) Jitter streaming. A frame captured via camera with slight shaking as it approaches the balloon. It can be seen that there is a significant brightness change in the field of view. (b) For scenes where camera is fixed and balloon is approaching camera screen. (c) For the two scenarios (a) and (b), the average brightness over time is displayed in (c), showing that the grayscale over time of SS is stable, while that of JS fluctuates dramatically. (d) The grayscale variations between adjacent frames under two different streaming are exhibited in (d), and it is found that the edge changes of objects are fuzzy in the scene of jitter streaming. (e) The summation and variance regions of grayscale changes along time between two adjacent frames are shown on the (e). This figure again illustrates the dramatic fluctuations of grey value variation in JS, distracting the extraction of visual looming features. The solid lines in red and blue are the grey change curves after summation.

the robot is highly susceptible to jitter streaming caused by fluctuations in the robot's vision. When the camera is not stabilized by gimbals, it follows the shake of the robot's body, which may result in jitter streaming as the robot traverses uneven terrain or experiences external forces. Although the jitter streaming still captures real-time motion information from external scenes, the resulting camera shake significantly interferes with the effective extraction of visual motion data, rendering the existing LGMD model ineffective.

To address the impact of jitter streaming on LGMD-based models, we conducted a quantitative comparison between jitter streaming and stationary streaming, focusing on brightness variation and its influence on the transmission of visual looming information. Our analysis revealed that, in contrast to stationary streaming, jitter streaming induced significant variations in grey values across the field of view, as illustrated in Fig.1, accompanied by fluctuating brightness levels. Given that existing LGMD-based models integrate visual information to extract looming features from the entire visual field, the accumulation of pronounced fluctuations in brightness can pose a distraction to LGMD models, leading to frequent and potentially erroneous collision warnings. Consequently, visual inputs from jitter streaming detrimentally impact the extraction of visual approaching motion features by existing LGMD models, thereby compromising collision detection accuracy.

Encouragingly, evidence from neuroscience research suggests that feedback loops in the optic circuit play a crucial role in enabling the brain to generate a coherent visual perception of dynamic natural scenes, particularly in the presence of unsteady vision—a capability not inherently present in existing LGMD-based models. Specifically, electroencephalography (EEG) experiments have shown that

feedback-related neural activity encodes visual information about incoherent visual input. Based on the functional magnetic resonance imaging (fMRI) data, the scene-selective cortex and human middle temporal complex (hMT) are identified as likely sources of this feedback[15]. This underscores the notion that constructing coherent visual experiences from the spatially fragmented mosaic of visual signals relies on cortical feedback rhythms traversing the visual hierarchy. However, the specific feedback mechanisms that enhance coherent dynamic visual perception remain unknown despite these findings. Therefore, further exploration is needed to develop a set of feedback mechanisms applied to LGMD models, providing engineering solutions against jittery video streams.

To explore the specific feedback mechanism enhancing coherent visual perception, this study investigates the efficacy of incorporating dynamic temporal variance feedback loops regulated by scalable functional to improve LGMD performance in the context of jitter streaming. The proposed feedback mechanism extracts second-order dynamic temporal variance information from the output of higher-order neurons within the LGMD network to assess the fluctuation level of the local neural response. This dynamic temporal variance is then regulated by a scalable functional to discriminate the variance caused by temporally incoherent visual input. The regulated signal is fed back to lower layers in negative feedback loop, ultimately improving the coherence of the visual input. To systematically substantiate the effectiveness of the proposed feedback mechanism, we conducted experiments on both stationary and jitter videos. The results indicate that, for jitter videos, the LGMD model integrated with our proposed feedback mechanism adeptly extracts collision detection characteristics and exhibits a notable resilience against Gaussian noise and salt-and-pepper noise. The main contributions of this paper can be summarized as follows:

- We introduce a novel feedback mechanism into the traditional LGMD-based neural network, effectively promoting its performance against jitter streaming.
- We systematically collected a jitter streaming dataset containing objects approaching in multiple scenes to validate the effectiveness of looming perception against jittered vision.
- This study offers a novel insight into the concrete ways in which a feedback loop promotes coherent visual perception, highlighting the potential of feedback mechanisms in enhancing visual neural abilities.

The rest of this paper is organised as follows: Section 2 provides an overview of feedback related research. Section 3 presents the proposed feedback LGMD. Section 4 reports the experimental results and performance comparisons with the classical LGMD model on real datasets. Finally, Part 5 summarizes this paper and points out the possible direction of future work.

2. Related Work

In this section, we briefly review feedback calculations in signal processing and then describe multiresolution analysis and total variance.

2.1. Feedback Computation in Signal Processing

In mammals, in addition to bottom-up feedforward neural networks, the visual cortex receives a dense feedback input network from higher-order cortical areas[16]. The information transmission of the feedback loop to the lateral geniculate nucleus(LGN)[17] has been demonstrated. It has been found that about five to ten per cent of the input from the lateral geniculate nucleus in the cat's visual system comes from the retina, while the feedback connections from the sixth layer cells of the visual cortex provide about thirty per cent of the synaptic input. Interestingly, in studying primates, the feedback projection was found to be almost equal in density to the feedforward neurons that flow through ventral vision[18, 19]. Generally, in the visual system, it is believed that the higher the hierarchy level, the longer it takes for information generated by visual stimulus to be transmitted[20, 18, 21]. There was also a delay in the response of neurons with feedback channels connecting lower visible areas from top to bottom[22]. In addition, more and more research shows that feedback signals experience slow time oscillations[23, 16]. Therefore, feedforward and feedback interactions between different cortical regions involved in sensory processing must be coordinated to generate comprehensive perception in the behavioural environment. In the process of information processing by the feedforward neural network, the feedback loop is activated after a certain time delay. But the mechanism behind these different feedback features is unclear. In the visual system of insects, the secondary activation of lower-level neural units with temporary delay caused by feedback has also been found. such as drosophila[24] and locust[25, 26]. At present, there is almost no modelling of locust vision system.

The feedback connection of these local loops of mammals also provides a way to affect collision detection. One immediate question is, how does a feedback loop function in a collision detection system, a possibility we will provide in this paper.

2.2. Multiresolution Analysis

Multiresolution analysis is one of the important concepts in wavelet analysis. Wavelet analysis projects data onto a set of basis functions to separate different scale information. Multiresolution analysis is the theoretical basis of the transform[27, 28, 29]. A direct application of multiresolution analysis is to construct an orthonormal wavelet basis for a discrete wavelet transform algorithm, which approximates data of different resolutions by projecting the data orthographically into different spaces $\{V_j\}_{j\in\mathbb{Z}}$ in $L^2(\mathbb{R})$($\mathbb{Z}$ is the set of all integers, $\mathbb{R}$ is the set of all real numbers, j,k$\in$ $\mathbb{Z}$). To compute these projections,

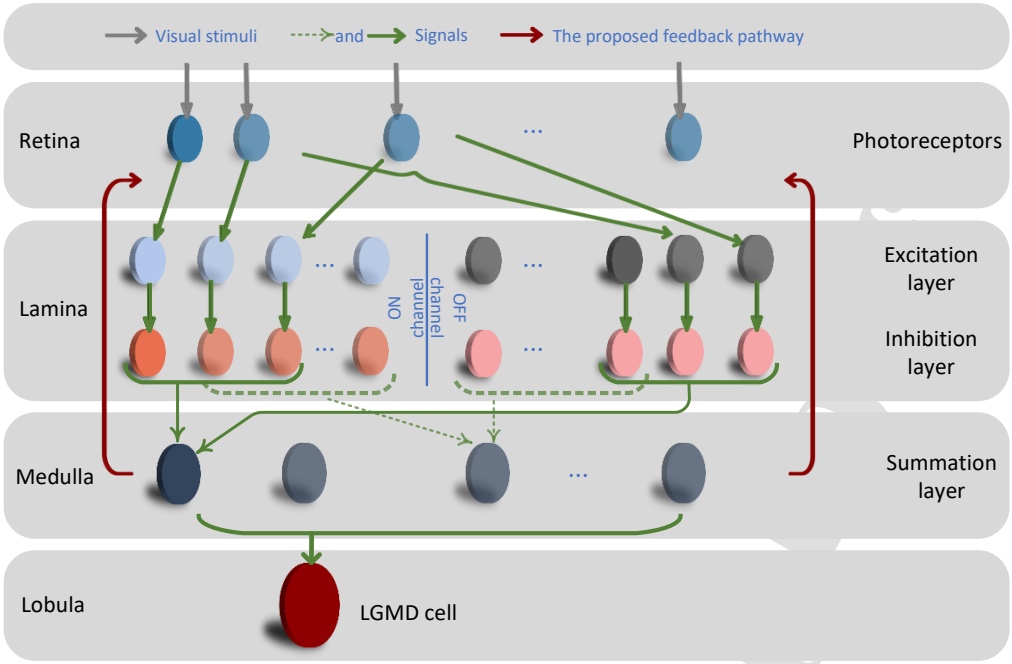

Figure 2: Overall flow chart of feedback LGMD neural network is presented. It consists of four nerve layers: retina, lamina, medulla, lobule, and feedback loops (solid red line).

it is necessary to give an orthonormal basis for V_j, constructed by dilating and translating a scaling function ϕ, denoted by $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$. Let space W_j is complement of the space V_j in V_{j+1}. The orthonormal basis for the space W_j is constructed by dilating and translating a single scalable function ψ called the mother wavelet function, whose translated and dilated versions are expressed by $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$.

A typical exemplar of scalable function is the spline function. An mth-order spline function is composed of piecewise mth-degree polynomials, implying that it possesses piece-wise (m-1)th-order differentiability [30, 31]. The m-th order cardinal B-spline $N_m(x)$ can be derived through the following recursive convolution, that is

$$N_m(x) = \int_{-\infty}^{+\infty} N_{m-1}N_1(t)\,dt = \int_0^1 N_{m-1}(x-t)\,dt, \quad (1)$$

$$N_1(x) = \begin{cases} 1 & \text{for } 0 \le x < 1 \\ 0 & \text{otherwise}\ , \end{cases} \quad (2)$$

where $N_1(X)$ denotes the Haar scaling function. In discrete form, the m-th order cardinal B-spline can be expressed as $N_m(x) = 2^{-m+1}\sum_{k=0}^{m}\binom{m}{k}N_m(2x-k)$, where $0 \le k < m$.

This paper focuses on the radial B-spline scalable function due to its simple structure and its successful application in wavelet neural networks. To obtain the radial B-spine scalable function, the B-spline N_m is introduced a $m/2$ shift along the x-axis. Then, the corresponding multi-dimensional frame is established by the translating radial B-spline scaling functions, which can be formulated by,

$$\Phi_{j_0,k}(x) = 2^{\frac{nj_0}{2}} N_m\big(\|2^{j_0}(x-k)\| + \frac{m}{2}\big). \quad (3)$$

Here, $n \in \mathbb{Z}$ represents the dimensionality of the input data x, $j_0 \in \mathbb{Z}$ denotes the scale parameter, and $k \in \mathbb{Z}^n$ signifies a vector of translation parameters.

In this paper, the dynamic signal within the LGMD network is treated as a function. Consequently, the multi-scale analysis tool operates not on a single point within a function, but on the function as a whole. Therefore, when constructing the feedback pathway for the LGMD network, the multi-scale processing of the feedback signal is approached from a functional perspective.

2.3. Total Variance

The total variance norm, serving as a type of norm, is frequently utilized in regularizing the image-based optimization problem including image denoising [32], image recovering [33], optical flow [34], and the like. The total variance norm can evaluate the level of variance accumulation of signals to characterize signal fluctuations, making a well-matched indicator to express the coherence of the signal. In existing image-based problems, the total variance norm is treated as a spatial operator to assess the spatial coherence of the pixel value [35]. Mathematically, the total variance norm can be calculated by

$$\|f\|_{TV} = \sum_{x,y}\sqrt{\big(\frac{\partial}{\partial x}f(x,y)\big)^2 + \big(\frac{\partial}{\partial y}f(x,y)\big)^2}, \quad (4)$$

where f denotes the pixel value of the image, $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ denote the spatial partial derivative operator.

In the feedback pathway developed in this paper, we modify the total variance norm into a temporal operator with a sliding accumulation window to evaluate the temporal coherence of the signal in the LGMD network at each time step.

3. Methods and Formulations

In this section, the feedback LGMD model is introduced. The proposed visual system model is shown in Fig.2, which is composed of feedforward and feedback pathways, wherein the feedforward pathway includes retina, lamina, medulla and lobule. These neurons coordinate with each other to detect potential collision threats in dynamic complex environments. As shown in Figure 2, the brightness signal is received by the retinal layer and then flows into the excitation and inhibition layers to calculate the change of the signal over time. The outgoing information is received by the feedback loop and again into the retina layer, and finally, all the information is integrated into the LGMD unit to identify potential collisions. More details are described below:

3.1. Retina Layer

In the retina, photoreceptors are arranged in a matrix to capture the grayscale brightness. The photoreceptors receive stimuli from the external environment and further extract changes in brightness over time. Let $I(x, y, t) \in \mathbb{R}^3$ represent the input image stream, where x and y represent the spatial position and t represent the current frame. We use $P(x, y, t)$ to describe this change, which is formally expressed as follows

$$P(x, y, t) = \int \big(\delta(t - \tau) - \delta(t - \Delta t - \tau)\big) I(x, y, \tau) d\tau, \quad (5)$$

where δ is the unit impulse function.

3.2. Lamina Layer

3.2.1. Feedback Pathway

As shown in Fig. 3, the proposed B-spline feedback signals are parallelly integrated into the input of ON/OFF channels. Specifically, the $P(x, y, t)$ signal and the feedback signals are both transmitted to separate ON/OFF channels. Then, the half-wave rectifier of ON channel allows positive part of $P(x, y, t)$ while blocking negative part, in contrast to the half-wave rectifier in OFF channel, which blocks positive part. Here, half-wave rectifiers receive not only information from the output of photoreceptor layer but also feedback signal from feedback loop, which can be formulated as follow:

$$P_{on}(x, y, t) = [P(x, y, t) + \beta_1 \Phi(x, y, t)]^+, \quad (6)$$

$$P_{off}(x, y, t) = [P(x, y, t) + \beta_2 \Phi(x, y, t)]^-, \quad (7)$$

where $[f]^+$ refers to max(x, 0) and $[f]^-$ denote max(-x, 0). $\Phi(x, y, t)$ is feedback function, which will be given in formula 19.

3.2.2. ON Channel

After that, the signal is directly mapped to the excited(E) layer without time delay. The calculation definition is as follows

$$E_{on}(x, y, t) = \iint P_{on}(u, v, t) G_{\sigma_1}(x - u, y - v)\, du\, dv, \quad (8)$$

where $E_{on}(x, y, t)$ represents excitation signal of ON channel, and $G_{\sigma_i}(x, y)(i = 1, 2)$ denotes Gaussian function, that is

$$G(x, y) = \frac{1}{2\pi\sigma_i^2} e^{(-\frac{x^2+y^2}{2\sigma_i^2})}, \quad (9)$$

$\sigma_i (i = 1, 2)$ is the standard deviation.

The information from the P layer is delayed to inhibition(I) layer of ON channel, and the $I_{on}(x, y, t)$ signal diffuses to surrounding area, forming lateral inhibition to affect the motion excitation. We have the output of $I_{on}(x, y, t)$, that is

$$I_{on}(x, y, t) = \iiint P_{on}(u, v, \tau)\varphi(t-\tau) G_{\sigma_2}(x-u, y-v) du dv d\tau, \quad (10)$$

where $G_{\sigma_2}(x, y)$ is Gaussian function, $\varphi(t)$ is temporal delay filter, which satisfies the following two properties

1) for $t \in [0, \infty)$,

$$\int \varphi(t) dt = 1, \quad (11)$$

2) $\varphi(t)$ is nonincreasing.

In this paper, a function satisfying these two conditions is applied, that is

$$\varphi(t) = 2e^{-\pi t^2}, \quad (12)$$

After that, in the summation(S) layer of ON channel, the information of excitation and inhibition layer are integrated, and it is obtained by linear computation, that is

$$S_{on}(x, y, t) = [E_{on}(x, y, t) - \alpha \cdot I_{on}(x, y, t)]^+, \quad (13)$$

where $S_{on}(x, y, t)$ denotes Local summation signal, α is coefficient to control the intensity of inhibitory flows.

3.2.3. OFF Channel

In the OFF channel, the local excitation signal is obtained without time delay, and the local inhibition signal is formed by delaying and convolving surrounding P_{off} signal. The whole process can be described as

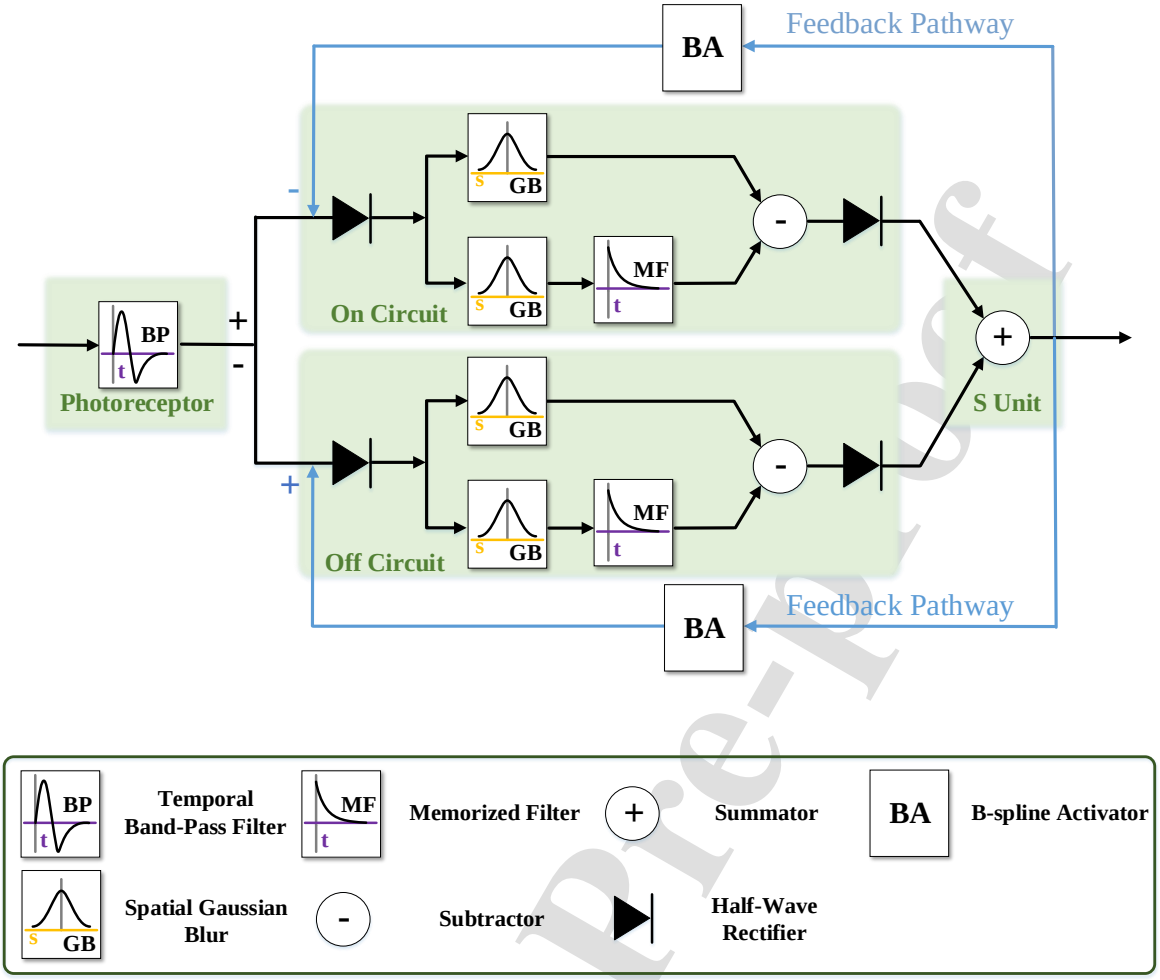

Figure 3: The schematic diagram of our proposed feedback LGMD model. External visual stimuli feed into photosensitive layer, the processed signal enters ON/OFF channels, ultimately flowing into summation layer. The feedback loop is reversely transferred from summation layer in the form of a radial B-spline scale function to ON/OFF channels.

$$E_{off}(x,y,t) = \iint P_{off}(u,v,t) G_{\sigma_1}(x-u, y-v)\, du\, dv, \tag{14}$$

$$I_{off}(x,y,t) = \iiint P_{off}(u,v,\tau)\varphi(t-\tau) \times G_{\sigma_2}(x-u, y-v) du dv d\tau, \tag{15}$$

where $E_{off}(x,y,t)$ and $I_{off}(x,y,t)$ denote excitation and inhibition signal of ON channel, respectively. $G_{\sigma_i}(x,y)(i=1,2)$ represents Gaussian function.

As shown in Fig. 3, excitation and inhibition signals flow into the summation layer, and signals are linearly integrated. α is the inhibition weight to adjust the contribution level of excitation signal and inhibition signal. The summation signal is denoted by $S_{off}(x,y,t)$, defined as follows

$$S_{off}(x,y,t) = [E_{off}(x,y,t) - \alpha \cdot I_{off}(x,y,t)]^+, \tag{16}$$

where α is inhibition strength coefficient.

3.3. Medulla Layer

3.3.1. Summation unit

The summation(S) unit integrates signals from the ON and off channels. The summation signal can be formulated as

$$S(x,y,t) = \varepsilon_1 \cdot S_{on}(x,y,t) + \varepsilon_2 \cdot S_{off}(x,y,t), \tag{17}$$

where $\varepsilon_i (i=1,2)$ refers to the combination of term coefficients.

3.3.2. The Proposed Feedback Loop

From Fig. 2, it can be seen that after obtaining information of summation layer, the signal not only flows

Table 1: Parameters of feedback LGMD Model

Equation	Parameters	Description
(6)	$\beta_1 = -0.01$	feedback coefficient of ON channel
(7)	$\beta_2 = 0.003$	feedback coefficient of OFF channel
(8)	$\sigma_1 = 0.1$	standard deviation of the Gaussian function in ON/OFF channels
(10)	$\sigma_2 = 0.3$	standard deviation of the Gaussian function in ON/OFF channels
(13)	$\alpha = 0.1$	inhibition coefficient
(14)	$\sigma_1 = 0.1$	standard deviation of the Gaussian function in ON/OFF channels
(15)	$\sigma_2 = 0.3$	standard deviation of the Gaussian function in ON/OFF channels
(16)	$\alpha = 0.1$	inhibition coefficient
(17)	$\varepsilon_1 = 0.2$	term coefficient in S later
(17)	$\varepsilon_2 = 0.2$	term coefficient in S later
(19)	$\varrho = 1$	scale parameter
(19)	$\eta = [100, 300]$	translation coefficient
(19)	$\zeta = [0.5, 1]$	spline coefficient

into LGMD, but also forms feedback signal that streams into feedback loops. The proposed feedback mechanism first extract the dynamic temporal variance information, which can be calculate by utilizing total variance norm as described as follow,

$$\begin{aligned} & D(x,y,t) = \|S(s,y,t)\|_{TV} \\ = & \int_{[0,T]} \Big| \int \big(\delta(\tau) - \delta(\tau - \Delta t)\big) S(x,y,t-\tau-s) d\tau \Big| ds. \end{aligned} \tag{18}$$

The extracted variance information is then regulated by a scalale functional. Here we utilize the B-spline functional as an exemplar due to its simple structure. The regulation can be formulated by

$$\Phi(x,y,t) = \zeta N_m\Big(\varrho D(x,y,t) - \eta K(x,y,t)\Big), \tag{19}$$

where $\Phi(x,y,t)$ denotes feedback function. ζ is the dilation or scale parameter, $K(x,y,t)$ valuing in $\mathbb{Z}$ represents the translation parameter. The mth-order cardinal B-spline $N_m(x)$ can be obtained by using Formula 1 and 2.

3.4. lobula Layer

In the lobule, LGMD cells integrate all the presynaptic stimuli from the S unit (see Fig. 3), resulting in the following membrane potential(MP)

$$K(t) = \frac{1}{1 + e^{-\int_1^R \int_1^C S(x,y,t) dx dy \cdot (C \cdot R \cdot \mu)^{-1}}}, \tag{20}$$

where C and R represent the columns and rows of the field of vision. μ denotes scale coefficient. And the output is in the range [0.5, 1).

4. Experimental results and analysis

To verify the validity of the feedback lgmd model, the proposed model is compared with LGMD1[10], LGMD2[36], F(ONn)-LGMD[37] and F(OFFn)-LGMD[37].

4.1. Experimental Setup

4.1.1. Parameters of the System

Based on our experiments, Table 1 lists the proposed model parameter Settings. The feedback coefficient is to satisfy the corresponding function of the model. According to experience, we take β_1=0.001, β_2=-0.01. Note that the values of the parameters R and C in the model are determined according to the length and width of the image sequence used. Unless otherwise stated, other parameters will not be changed in subsequent experiments.

4.1.2. Datasets

The proposed feedback LGMD model is evaluated using real datasets. The datasets include: a) in natural environment, a GoPro camera mounted on a tripod recorded the balloon approaching camera and colliding, please note that the camera is stationary at this time. b) Hang the balloon on a wooden pole, camera keeps shaking as it approaches the object. c) In the natural environment, the balloon and the camera come close to each other and collide. At this time, the camera is constantly shaking. d) The resultant dataset is sequences of images in a) that have been corrupted with Gaussian noise and salt and pepper noise. e) In reality, a sequence of images capturing car in motion. Among them, the resolution of image sequences captured from a)-c) is 270 × 480 pixels, with a frame rate of 30 frames per second (fps). The video recorded in e) includes scenes of cars colliding and cars not colliding, and the captured real image sequence has resolution of 240 × 426 and 240 × 432[38]. It is worth mentioning that the videos we shot are not limited to balloons, including collision scenes with bicycles, pedestrians, buildings, etc.

4.1.3. Evaluation Criteria

We use the Interframe Similarity Index (ISI) based on the Structural Similarity Index(SSIM)[39, 40, 41, 42] to evaluate stabilization of video. The ISI is the average of the similarity between consecutive frames in a video. The

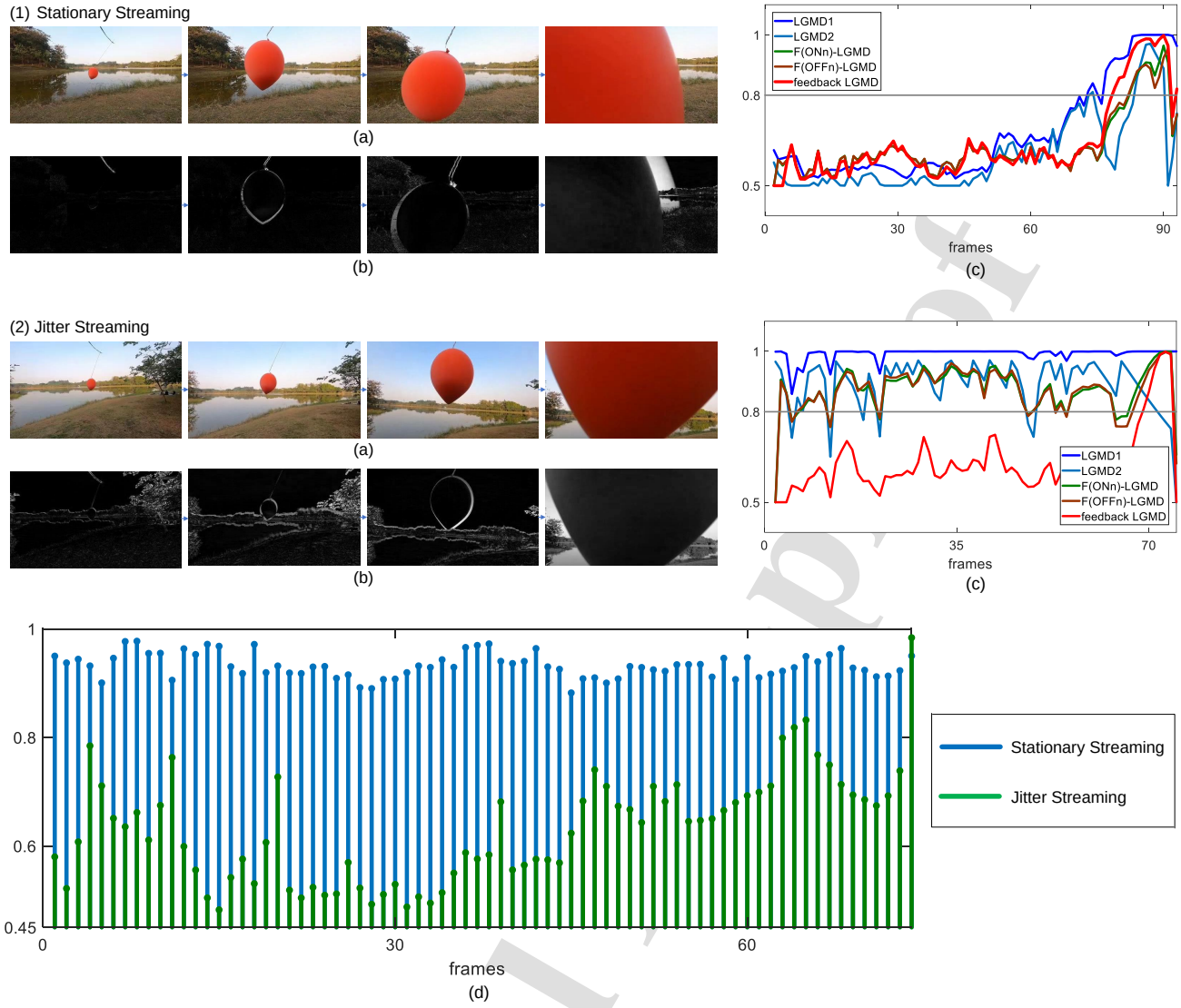

Figure 4: (1) For the fixed camera, scenes where balloon approaches camera and collides. (2) The balloon is suspended and camera is manually held to slam balloon. (d) shows the structural similarity index(SSIM). According to distribution of SSIM, the value of SSIM of Stationary Streaming is close to 1 and changes little, while the value of SSIM of jitter streaming changes frequently. In (c), the ordinate is the membrane potential.

Video Stabilization Quality Assessment(VSQA) metric is described as

$$ISI = \frac{1}{n-1}\sum_{i=1}^{n-1} SSIM(t), \tag{21}$$

Where SSIM(t) represents the structural similarity index between the i-th frame and the (i+1)th frame. The larger the ISI value, the smaller the variation between consecutive frames, which indicates a more stable video.

We use the Membrane Potential Stability Index(MpSI) to evaluate the collision detection performance of the model, it can be described as

$$MpSI = \frac{max(Mp)}{mean(Mp)}, \tag{22}$$

where max(Mp) denotes the maximum value of membrane potential, mean(Mp) represents the average membrane potential. When membrane potential exceeds the threshold, a larger MpSI value indicates a better collision detection ability of the model.

We propose to measure the sensitivity of the model using the Sensitivity Index (SI), which is obtained by dividing the first frame(F_n) above threshold by total number of frames(N_f). SI is described as

$$SI = \frac{F_n}{N_f}. \tag{23}$$

Obviously, the earlier membrane potential exceeds the threshold, the smaller the value of SI.

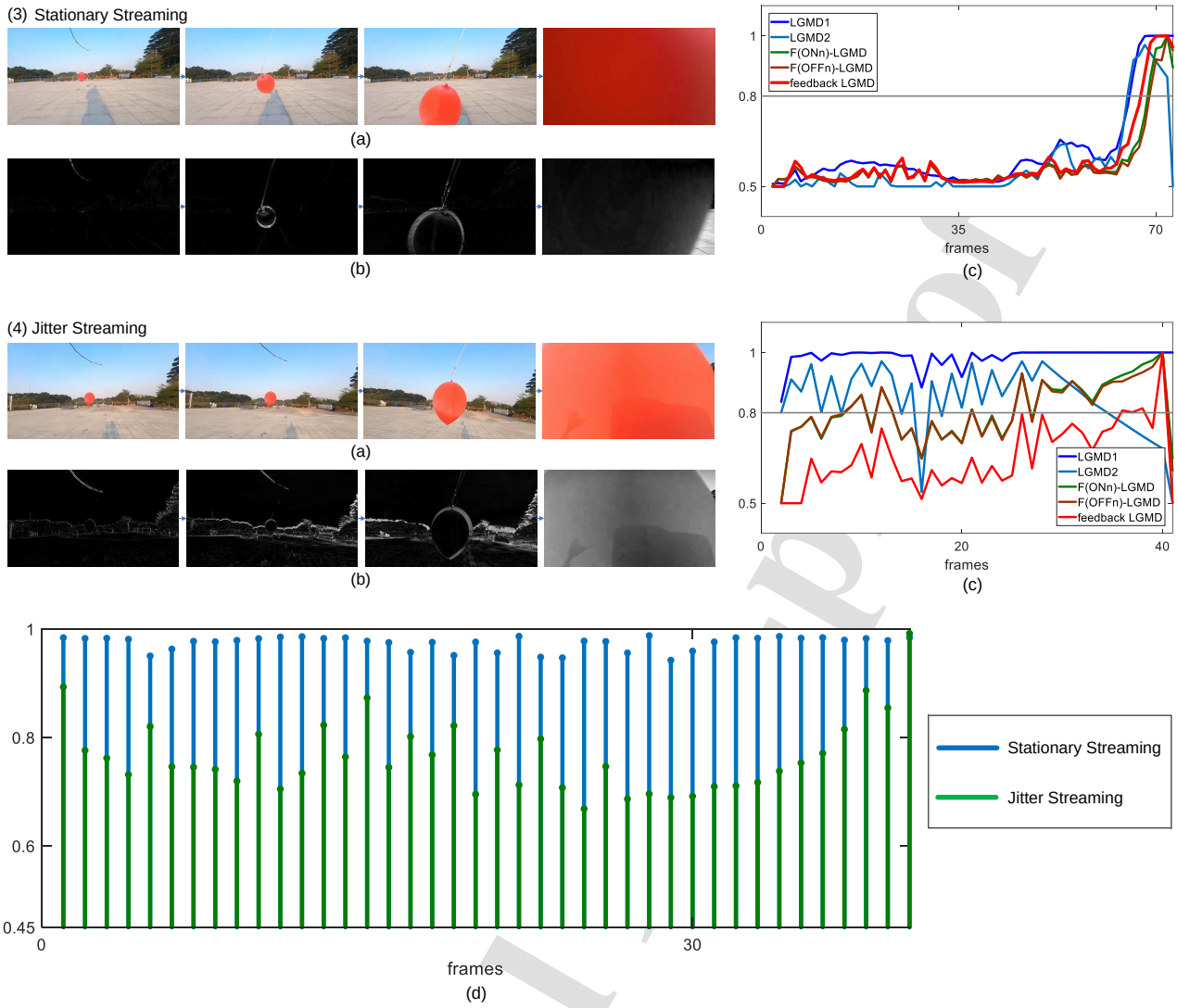

Figure 5: (3) is sequence of images taken by fixed camera. (4) is scenario in which balloon is suspended and camera is manually held to slam balloon. From the visualization of P layer in (3)(b) and (4)(b), it can be clearly seen that the amount of change caused via background noise of the jitter streaming is larger than that of (3). According to distribution of SSIM of (d), it is obvious that SSIM distribution of jitter streaming is chaotic, while SSIM value of stationary streaming is close to 1.

4.2. Effectiveness of Model to Collision Threat

We compare membrane potential output of feedback LGMD model with other LGMD models and report the effective recognition of feedback LGMD model for unstable videos. To validate the effectiveness of proposed feedback LGMD model in collision detection, we first test it using image sequences captured by fixed camera, Fig. 4 (1) and Fig. 5 (3) show experimental results. To better illustrate the extent of image changes within scene, we performe visualization processing on the P layer(feedback LGMD and other LGMD models undergo the same P-layer processing procedure). As a result, it can be clearly seen that the changes of P layer are not very pronounced. From the distribution of SSIM values in Fig. 4 (d) and Fig. 5 (d), it is evident that both scenarios exhibit relative stability. Furthermore, as shown in Table 2, the image sequences of objects approaching fixed camera are also relatively stable, with ISI values of 0.9057 for Fig. 4 (1) and 0.9595 for Fig. 4 (3), respectively. These values are relatively large, close to 1, indicating that the similarity between image frames is significant. From Fig. 4(1) (c) and Fig. 5 (3) (c), it can be observed that membrane potential curves for feedback LGMD, LGMD1, LGMD2, F(ONn)-LGMD and F(OFFp)-LGMD all gradually increase, with membrane potentials exceeding the threshold of 0.8 before collision occurs. In Table 2, for Fig. 4 (1) and Fig. 5 (3), the MpSI values of feedback LGMD, LGMD1, LGMD2, F(ONn)-LGMD and F(OFFp)-LGMD models all exceed 1.5. It indicates that five models perform well in collision detection for camera-fixed scenes. As can see from Table 3, for Fig. 4 (1) and Fig. 5 (3), the SI for all models is larger, greater than 0.7.

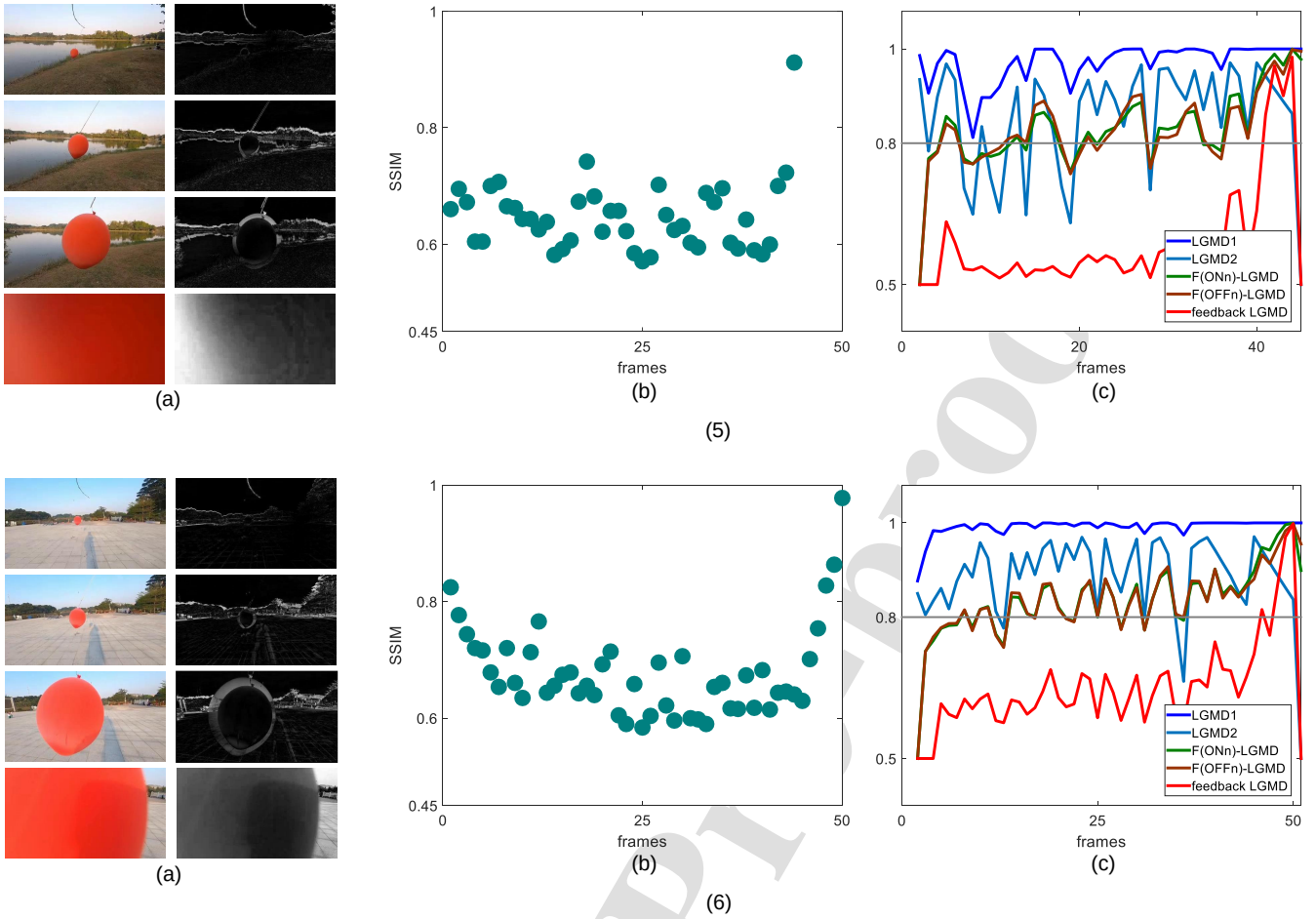

Figure 6: In scenarios (5)(a) and (6)(a), the relative motion between balloon and camera is depicted. From visualizations of P layer in (5)(b) and (6)(b), it is evident that there is a significant amount of variance introduced via background noise. According to the distribution of SSIM of (5)(c) and (6)(c), comparing Fig.4(1), it is found that the image sequence in Fig.6 is relatively unstable. Observing the membrane potential curves of feedback LGMD and other models, similar to Fig.4(2), the membrane potential of feedback LGMD model gradually exceeds threshold.

4.3. Effectiveness of Model for Jitter Videos

After that, to reveal the response characteristics of feedback LGMD neural network, we report the membrane potential output of feedback LGMD and other LGMD models in the scenario where object is stationary, while camera is constantly shaking and approaching object. The jitter mentioned here indicates stable deterioration of image sequences, implying that the ISI value is relatively small (see Table 2). For videos take in two different locations, namely Fig. 4 (1) and Fig. 5 (3) (Fig. 4 (2) and Fig.5 (4)), it can be seen from the visualization of P layer that the amount of change in P layer of jitter streaming is more than that in Fig. 4 (1) and Fig. 5 (3). ISI values for Fig. 4 (2) and Fig. 5 (4) are 0.6313 and 0.7648, respectively, both of which are smaller than ISI values of video in Fig. 4 (1). Compared with Fig.4 (1) and Fig. 5 (3), the distribution of SSIM in Fig. 4 (2) and Fig. 5 (4) are more erratic, indicating that the scene in Fig. 4 (2) and Fig. 5 (4) is unstable. Fig. 4 (2) (c) and Fig.5 (4) (c) display the membrane potential curves of five models to two scenario stimuli. It can be clearly seen that the membrane potential response of other models are already saturated, while the membrane potential curve of feedback LGMD model gradually increases and effectively identifies collision threats. Table 2 also shows MpSI of all models. For Fig. 4 (2) and Fig. 5 (4), the MpSI of the feedback LGMD model are 1.5605 and 1.5212, respectively, while the MpSI of other models is less than 1.5. These values also show that feedback LGMD model has a favourable response to collision threat. In the above experiment, feedback loop demonstrates its ability to improve performance of the model against collision threats by suppressing background noise caused by 'jitter'. Relative motion is considered a vital clue affecting dynamic response characteristics of neurons. From this perspective, the proposed feedback LGMD provides a possible explanation for promoting object recognition.

For scenes where both object and camera are moving,

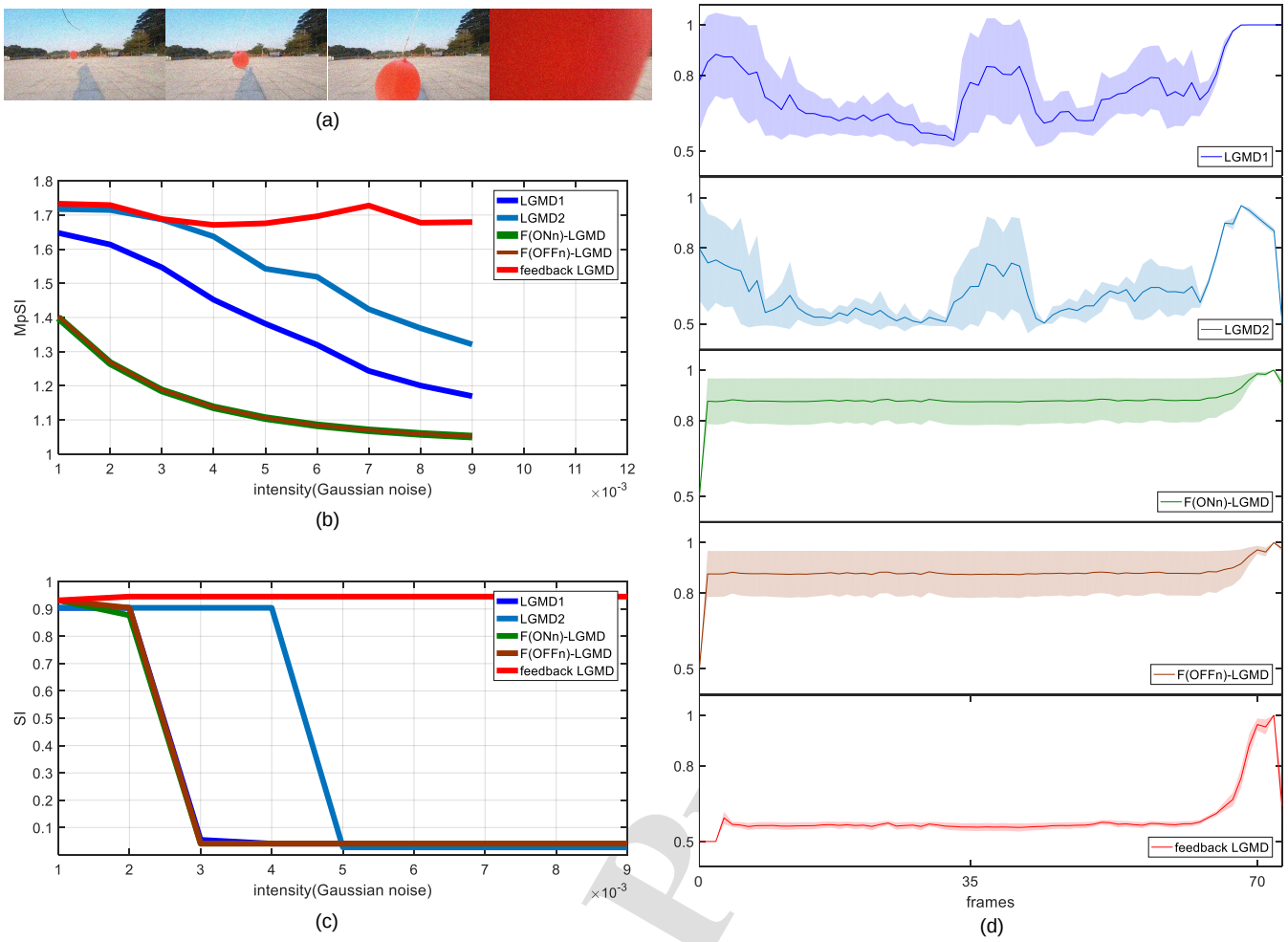

Figure 7: For the scenario in Fig.5, Gaussian noise with variances ranging from 0.001 to 0.009 is added. For scenario (a), the MpSI of other models is gradually decreasing, the overall MpSI of feedback LGMD is higher than that of other models. The membrane potential variance statistics of five models also show that feedback LGMD model is less affected by noise. The SI of feedback LGMD change little overall.

Fig. 6 shows the membrane potential response of feedback LGMD and other LGMD models to these scenes. It can be seen from Fig. 6 (5) and (6) that compared with Fig. 4 (1), the output variation of P layer is more, but compared with Fig. 5 (4), the variation of P layer is difficult to evaluate. The relatively strong jitter in these two scenes is also evident from the distributions of SSIM in Fig. 6 (5) (c) and (6) (c). In Table 2, the ISI values of Fig. 6 (5) and (6) are both less than 0.7, indicating that videos are also unstable. Fig.6 (5) (d) and (6) (d) show that the membrane potential of other models quickly reach saturation due to the "jitter" during the video shooting process, and feedback LGMD model can effectively identify collision threats. The MpSI of LGMD model in Table 2 is also approaching 1, which is a poor performance.

We have verified that unstable videos cannot elicit strong responses from the proposed model, as feedback mechanism in the model suppresses stimulus generated by visual scene changes.

4.4. Effectiveness of Model for Noise Scenario

In LGMD1 and LGMD2 model, G layer is added after S layer, which realizes the response to coherent stimulus and weakens stimulus of the small noise generated by background. We test the proposed model by adding different Gaussian and salt-and-pepper noise levels to image sequences in Fig. 5. Fig. 7 (a) shows the different Gaussian noise variances taken in the experiment (σ =0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009). Fig. 7 shows statistical plot of membrane potential variance after adding Gaussian noise to the scenarios, as well as the changes in MpSI of five models with variation of Gaussian noise variance.

As can be seen from Fig. 7 (b) that with the increase of noise variance, MpSI of feedback LGMD fluctuates, while MpSI of other models gradually decreases, indicating that Gaussian noise has an impact on the membrane potential output of four models. In Fig. 7 (c), SI of feedback LGMD model changes little with the increase of noise variance, and other models are greatly affected by noise In Fig. 7

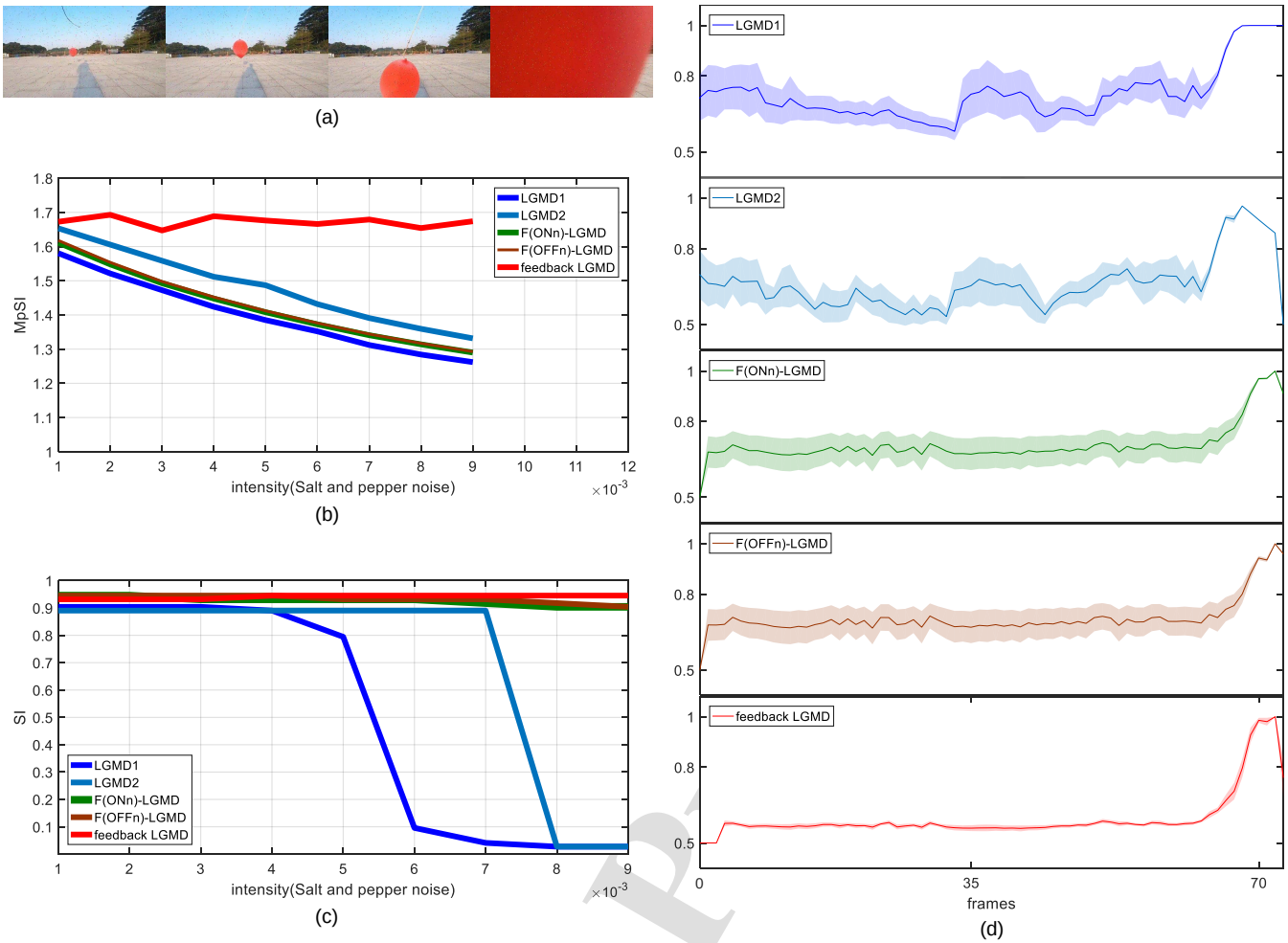

Figure 8: Salt and pepper noise is added to the scene in Fig.5, and the intensity of noise ranges from 0.001 to 0.009. (b) shows that MpSI of feedback LGMD is greater than that of other models and the statistical plots of membrane potential variance of five models also show that feedback LGMD model is less affected by noise.

(d), the film potential variance statistics of five models are shown, indicating that membrance potential output of feedback LGMD model is relatively stable.

Fig. 8 (a) shows the different noise intensities of salt and pepper noise taken in experiment (σ_1 =0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009). Similar to Fig. 7, it can be seen from Fig. 8 (b) that MpSI of other models gradually decreases with the increase of noise variance. However, overall, MpSI of feedback LGMD model is larger than that of other models. Fig. 8 (d) shows the membrane potential variance statistics of five models. It can be seen from these figures that membrane potential curve of other models fluctuate greatly.

Through the above two experiments, we realized that the model cannot provide feedback information to filter out noise in the background completely. However, compared to other models, the feedback loop weakens the impact of noise on the model in a short period. On this basis, it can effectively maintain potential collision threats.

4.5. Real-world stimuli testing

Finally, we test the proposed model with videos of car driving. The experiment adopts scene that: (1) a car collides in the process of driving, corresponding to Fig. 9 (a). (2) a car driving on road without any collision, corresponding to Fig. 9 (b). While driving a car, the intensity of sunlight exposure is not constant due to obstruction of trees, and surrounding background noise is diverse. The experimental results are shown in Fig. 9. As can be seen from Fig. 9 (a), the membrane potential of feedback LGMD model shows slow accumulation, while the membrane potential of LGMD1 model quickly exceeds the threshold. In the scenario where the car does not collide, the membrane potential of feedback LGMD model does not exceed the threshold value.

5. Conclusion and discussion

In this paper, we introduce a dynamic temporal variance feedback loop regulated by scalable functional into

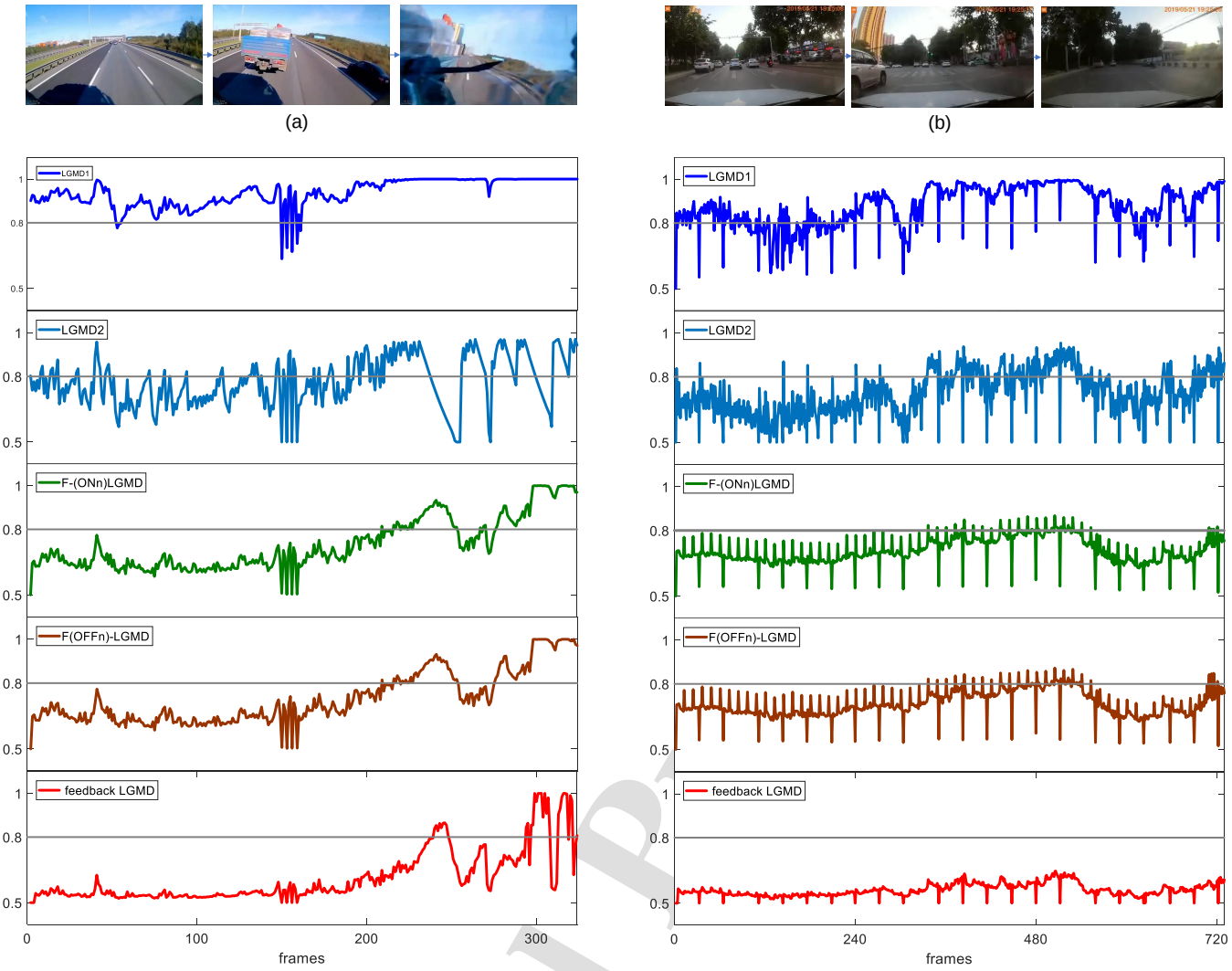

Figure 9: For ground vehicle scenarios: a) the car crashes, b) a car is driving on the road without any collision. Upon observing membrane potential curves of five models, it is noticed that feedback LGMD model is capable of accurately recognizing the collision.

the LGMD-based visual neural network for collision detection to promote its performance against the unstable visual input due to the jitter of the camera carrier during imaging. The computation architecture of this feedback model consists of two parts, including a feedforward neural network and a feedback loop. The feedforward network extracts motion information by integrating the brightness changes of adjacent pixels in two frames. In the feedback loop, the dynamic temporal variance information is extracted from the output of high-order neurons in the conventional network to assess the fluctuation by level of local neural responses and regulated by a scalable functional to differentiate variance induced by incoherent visual input. Then the regulated signal is reintegrated into the input through the negative feedback loop to reduce the effects of the jitter signal within the network. Numerical results on a collected jitter streaming dataset indicate that the proposed model can effectively resist the influence of jitter streaming. This study extends the limitations of the LGMD-based network against jitter streaming and also provides a novel insight into the potential feedback mechanism in facilitating the visual neural capabilities.

Indeed, jitter streaming is a common challenge in practical applications involving visual mobile robots, such as walking and wheeled robots. The body of a visual mobile robot can shake due to factors like uneven terrain, variations in posture during movement, and external environmental interference. This shaking compromises the stability of the visual sensors mounted on the robot, resulting in jitter streaming that undermines visual reliability, consequently, presenting substantial challenges for real-time visual motion detection, which is essential for understanding environmental dynamics. To address the issue caused by jitter streaming, one typical solution is to use gimbals to enhance visual stability. However, for microrobots, the constraints of physical size and energy consumption often

Table 2

Scenario	ISI	MpSI of feedback LGMD	MpSI of LGMD1	MpSI of LGMD2	MpSI of F(ONn)-LGMD	MpSI of F(OFFp)-LGMD
Fig.4 (1)	0.9057	1.5621	1.5047	1.6179	1.5360	1.5206
Fig.4 (2)	0.6313	1.5964	1.0216	1.0973	1.1519	1.1527
Fig.5 (3)	0.9595	1.7127	1.6508	1.7009	1.7523	1.7611
Fig.5 (4)	0.7648	1.5354	1.0399	1.1571	1.2515	1.2584
Fig.6 (5)	0.6477	1.6845	1.0522	1.1477	1.2144	1.2158
Fig.6 (6)	0.6772	1.5434	1.0286	1.1055	1.2062	1.2044

Table 3

Scenario	SI of feedback LGMD	SI of LGMD1	SI of LGMD2	SI of F(ONn)-LGMD	SI of F(OFFp)-LGMD
Fig.4 (1)	0.8495	0.7849	0.7957	0.8925	0.8925
Fig.4 (2)	0.9200	0.0267	0.0267	0.0400	0.0400
Fig.5 (3)	0.9315	0.9041	0.8904	0.9452	0.9452
Fig.5 (4)	0.8780	0.0488	0.0488	0.2195	0.2195
Fig.6 (5)	0.9111	0.0444	0.0444	0.1111	0.1111
Fig.6 (6)	0.9020	0.0392	0.0392	0.1569	0.1569

preclude the use of such stabilization devices. In recent decades, research on microrobots has gained increasing attention due to their convenience, adaptability to confined spaces, low power consumption, and cost-effectiveness, making them particularly valuable in fields such as swarm intelligence and automated inspection. Thus, enhancing visual perception capabilities for microrobots in the face of jitter streaming aligns with both application needs and advancements in visual intelligence for micro agents.

Since the solution to handle jitter streaming from stabilizing the camera is ineffective for micro robots, an alternative approach is to develop unique mechanisms to extract information despite jitter streaming. This paper proposes a novel feedback mechanism designed to eliminate the impact of jitter on visual collision detection, tailored specifically to the characteristics of visual collision detection tasks. Notably, other visual tasks—such as object recognition, visual navigation, and target tracking—could also be significantly affected by jitter streaming, highlighting valuable avenues for future research aimed at enhancing the visual intelligence of microrobots. Moving forward, we will investigate the potential applications of the proposed model and combined with adaptive security control system[43] for intelligent robots to deal with more complex and dynamic visual scenes.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was supported in part by the National Natural Science Foundation of China under grant 12031003 and in part by International Cooperation and Exange Project, NSFC-PSF, No.12211540710.

References

[1] M. A. Alalawi, Desert locust control management: Innovative technology on the front lines in the battle against locust, in: 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 2020, pp. 4–4.

[2] J. Zhao, H. Wang, N. Bellotto, C. Hu, J. Peng, S. Yue, Enhancing lgmd's looming selectivity for uav with spatial–temporal distributed presynaptic connections, IEEE Transactions on Neural Networks and Learning Systems 34 (5) (2023) 2539–2553.

[3] W. Zhang, Z. Wei, Y. Xu, L. Peng, C. Dong, Y. Jin, Q. Shi, Design of a small-scale locust-inspired robot capable of jump-gliding locomotion, in: 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2023, pp. 1–6.

[4] M. O'Shea, J. Williams, The anatomy and output connection of a locust visual interneurone; the lobular giant movement detector (lgmd) neurone, Journal of comparative physiology 91 (1974) 257–266.

[5] J. Sztarker, F. C. Rind, A look into the cockpit of the developing locust: looming detectors and predator avoidance, Developmental neurobiology 74 (11) (2014) 1078–1095.

[6] P. J. Simmons, F. C. Rind, Responses to object approach by a wide field visual neurone, the lgmd2 of the locust: characterization and image cues, Journal of Comparative Physiology A 180 (1997) 203–214.

[7] F. C. Rind, S. Wernitznig, P. Pölt, A. Zankel, D. Gütl, J. Sztarker, G. Leitinger, Two identified looming detectors in the locust: ubiquitous lateral connections among their inputs contribute to selective responses to looming objects, Scientific Reports 6 (1) (2016) 35525.

[8] G. Schlotterer, Response of the locust descending movement detector neuron to rapidly approaching and withdrawing visual stimuli, Canadian Journal of Zoology 55 (8) (1977) 1372–1376.

[9] F. C. Rind, D. Bramwell, Neural network based on the input organization of an identified neuron signaling impending collision, Journal of neurophysiology 75 (3) (1996) 967–985.

[10] S. Yue, F. C. Rind, Collision detection in complex dynamic scenes using an lgmd-based visual neural network with feature enhancement, IEEE transactions on neural networks 17 (3) (2006) 705–716.

[11] S. Yue, F. C. Rind, Redundant neural vision systems—competing for collision recognition roles, IEEE Transactions on Autonomous Mental Development 5 (2) (2013) 173–186.

[12] C. Hu, F. Arvin, C. Xiong, S. Yue, Bio-inspired embedded vision system for autonomous micro-robots: The lgmd case, IEEE Transactions on Cognitive and Developmental Systems 9 (3) (2016) 241–254.

[13] Q. Fu, C. Hu, T. Liu, 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), 2017, pp. 3996–4002.

[14] J. Zhao, X. Ma, Q. Fu, C. Hu, S. Yue, An lgmd based competitive collision avoidance strategy for uav, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, 2019, pp. 80–91.

[15] D. K. Chen Lixiang, Radoslaw M Cichy, Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision, Science advances 9 (45).

[16] L. Nurminen, S. Merlin, M. Bijanzadeh, F. Federer, A. Angelucci, Top-down feedback controls spatial summation and response amplitude in primate visual cortex, Nature communications 9 (1) (2018) 1–13.

[17] W. Wang, H. E. Jones, I. M. Andolina, T. E. Salt, A. M. Sillito, Functional alignment of feedback effects from visual cortex to thalamus, Nature neuroscience 9 (10) (2006) 1330–1336.

[18] D. J. Felleman, D. C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex., Cerebral cortex (New York, NY: 1991) 1 (1) (1991) 1–47.

[19] O. Sporns, J. D. Zwi, The small world of the cerebral cortex, Neuroinformatics 2 (2004) 145–162.

[20] T. Bachmann, A hidden ambiguity of the term "feedback" in its use as an explanatory mechanism for psychophysical visual phenomena, Frontiers in Psychology 5 (2014) 780.

[21] M. T. Schmolesky, Y. Wang, D. P. Hanes, K. G. Thompson, S. Leutgeb, J. D. Schall, A. G. Leventhal, Signal timing across the macaque visual system, Journal of neurophysiology 79 (6) (1998) 3272–3278.

[22] L. G. Nowak, J. Bullier, The timing of information transfer in the visual system, in: Extrastriate cortex in primates, 1997, pp. 205–241.

[23] M. S. Avendaño, C. Leidy, J. M. Pedraza, Tuning the range and stability of multiple phenotypic states with coupled positive–negative feedback loops, Nature communications 4 (1) (2013) 2605.

[24] L. Zheng, G. G. de Polavieja, V. Wolfram, M. H. Asyali, R. C. Hardie, M. Juusola, Feedback network controls photoreceptor output at the layer of first visual synapses in drosophila, The Journal of general physiology 127 (5) (2006) 495–510.

[25] S. Wernitznig, F. C. Rind, P. Pölt, A. Zankel, E. Pritz, D. Kolb, E. Bock, G. Leitinger, Synaptic connections of first-stage visual neurons in the locust schistocerca gregaria extend evolution of tetrad synapses back 200 million years, Journal of Comparative Neurology 523 (2) (2015) 298–312.

[26] S. Wernitznig, F. C. Rind, A. Zankel, E. Bock, D. Gütl, U. Hobusch, M. Nikolic, L. Pargger, E. Pritz, S. Radulović, et al., The complex synaptic pathways onto a looming-detector neuron revealed using serial block-face scanning electron microscopy, Journal of Comparative Neurology 530 (2) (2022) 518–536.

[27] M. Thuillard, A review of wavelet networks, wavenets, fuzzy wavenets and their applications, Advances in Computational Intelligence and Learning: Methods and Applications (2002) 43–60.

[28] M. Stephane, A wavelet tour of signal processing, 1999.

[29] J. P. Muszkats, S. A. Seminara, M. I. Troparevsky, Applications of Wavelet Multiresolution Analysis, 2021.

[30] M. A. Unser, Ten good reasons for using spline wavelets, in: Wavelet applications in signal and image processing V, Vol. 3169, 1997, pp. 422–431.

[31] M. Unser, A. Aldroubi, M. Eden, B-spline signal processing. i. theory, IEEE transactions on signal processing 41 (2) (1993) 821–833.

[32] X. Bresson, T. F. Chan, et al., Fast dual minimization of the vectorial total variation norm and applications to color image processing, Inverse problems and imaging 2 (4) (2008) 455–484.

[33] Y. Chen, S. Wang, Y. Zhou, Tensor nuclear norm-based low-rank approximation with total variation regularization, IEEE Journal of Selected Topics in Signal Processing 12 (6) (2018) 1364–1377.

[34] M. Drulea, S. Nedevschi, Total variation regularization of local-global optical flow, in: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2011, pp. 318–323.

[35] X. Han, J. Wu, L. Wang, Y. Chen, L. Senhadji, H. Shu, in: Linear total variation approximate regularized nuclear norm optimization for matrix completion, Vol. 2014, 2014, p. 765782.

[36] Q. Fu, C. Hu, J. Peng, F. C. Rind, S. Yue, A robust collision perception visual neural network with specific selectivity to darker objects, IEEE transactions on cybernetics 50 (12) (2019) 5074–5088.

[37] Z. Chang, Q. Fu, H. Chen, H. Li, J. Peng, A look into feedback neural computation upon collision selectivity, Neural Networks 166 (2023) 22–37.

[38] Q. Fu, H. Wang, J. Peng, S. Yue, Improved collision perception neuronal system model with adaptive inhibition mechanism and evolutionary learning, IEEE Access 8 (2020) 108896–108912.

[39] S. K. Sandhu, A. Agarwal, Summarizing videos by key frame extraction using ssim and other visual features, in: Proceedings of the Sixth International Conference on Computer and Communication Technology 2015, 2015, pp. 209–213.

[40] W. Guilluy, A. Beghdadi, L. Oudre, A performance evaluation framework for video stabilization methods, in: 2018 7th European Workshop on Visual Information Processing (EUVIP), 2018, pp. 1–6.

[41] M. S. Ito, E. Izquierdo, A dataset and evaluation framework for deep learning based video stabilization systems, in: 2019 IEEE Visual Communications and Image Processing (VCIP), 2019, pp. 1–4.

[42] M. R. e Souza, H. Pedrini, Combination of local feature detection methods for digital video stabilization, Signal, Image and Video Processing 12 (2018) 1513–1521.

[43] Y. Bi, T. Wang, J. Qiu, M. Li, C. Wei, L. Yuan, Adaptive decentralized finite-time fuzzy secure control for uncertain nonlinear cpss under deception attacks, IEEE Transactions on Fuzzy Systems 31 (8) (2023) 2568–2580.

Declaration of Interest Statement

We declare that we have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.