Bio-Inspired Small Target Motion Detection With Spatio-Temporal Feedback in Natural Scenes

Hongxin Wang, Zhiyan Zhong, Fang Lei, Jigen Peng, and Shigang Yue, *Senior Member, IEEE*

***Abstract*— Small moving objects at far distance always occupy only one or a few pixels in image and exhibit extremely limited visual features, which bring great challenges to motion detection. Highly evolved visual systems endow flying insects with remarkable ability to pursue tiny mates and prey, providing a good template to develop image processing method for small target motion detection. The insects' excellent sensitivity to small moving objects is believed to come from a class of specific neurons called small target motion detectors (STMDs). However, existing STMD-based methods often experience performance degradation when coping with complex natural scenes. In this paper, we propose a bio-inspired visual system with spatio-temporal feedback mechanism (called Spatio-Temporal Feedback STMD) to suppress false positive background movement while enhancing system responses to small targets. Specifically, the proposed visual system is composed of two complementary subnetworks and a feedback loop. The first subnetwork is designed to extract spatial and temporal movement patterns of cluttered background by neuronal ensemble coding. The second subnetwork is developed to capture small target motion information where its output and signal from the first subnetwork are integrated together via the feedback loop to filter out background false positives in a recurrent manner. Experimental results demonstrate that the proposed spatio-temporal feedback visual system is more competitive than existing methods in discriminating small moving targets from complex natural environments.**

***Index Terms*— Biologically inspired visual system, small target motion detection, natural scenes, spatio-temporal feedback, population coding.**

I. Introduction

AUTONOMOUS mobile robots with visual sensors onboard, such as drones, space probes, and unmanned underwater vehicles, have shown great potential in performing a wide range of challenging tasks, from navigation in uncontrolled environments to military reconnaissance without the need for guidance devices [2]. Artificial visual system capable of efficient and robust motion perception is critical for mobile robots to enable autonomic responses to surroundings that is always highly complex, dynamic, and abundant with visual motion. For example, early detection of objects with potential threats in the distance would help intelligent robots readily seize a first-mover advantage in competition or interaction. However, when a target is far distant or extremely small, it generally cover only one or a few pixels in image, appearing as a dim speckle without prominent appearance information. Fig. 1 gives an example of two small moving targets in the distance. Due to their limited sizes and few appearance cues, it is difficult even for humans to perceive such tiny objects.

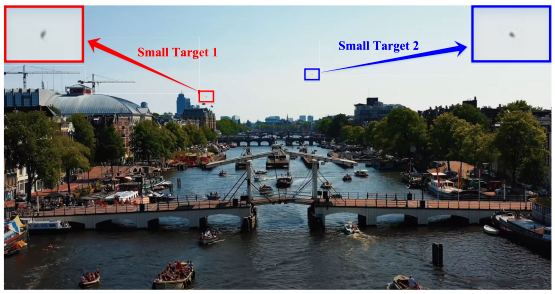


Fig. 1. Two small targets in the far distance that could be bird or drone [1]. The red-boxed and blue-boxed regions are enlarged, each of which contains a small target and its surrounding background. Due to long observation distance, the targets are all present as minute dim speckles in image, only a few pixels in size, with extremely limited visual features.

Small target motion detection against complex dynamic environment finds applications in a variety of domains including autonomous driving, military reconnaissance, and astronomic observation. However, it remains highly challenging to artificial visual systems, because 1) small targets are only one or a few pixels in size, exhibiting poor-quality appearance, let alone process discriminative visual features for motion detection; 2) small targets always display highly blurred boundaries and extremely low contrast, making them difficult to separate from numerous background clutters; 3) freely moving camera would bring further challenges for motion discrimination, such as strong parallax effects, constantly changing complex scene, and significant relative motion to objects of interest.

Great progress has been made towards motion detection of large objects with sufficiently detailed appearance and clearly defined structure from which rich visual features could be

Manuscript received 31 August 2023; revised 19 November 2023; accepted 11 December 2023. Date of publication 27 December 2023; date of current version 4 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 12031003 and Grant 62103112, in part by the European Union's Horizon 2020 Research and Innovation Program through the Marie Sklodowska-Curie under Grant 691154 STEP2DYNA and Grant 778062 ULTRACEPT, and in part by the Guangzhou Basic and Applied Basic Research Scheme under Grant 2023A04J0372 and Grant 2024A04J3271. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jun Liu. *(Hongxin Wang and Zhiyan Zhong contributed equally to this work.) (Corresponding authors: Shigang Yue; Jigen Peng.)*

Hongxin Wang and Jigen Peng are with the Machine Life and Intelligence Research Center, School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China (e-mail: jgpeng@gzhu.edu.cn).

Zhiyan Zhong is with the School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China.

Fang Lei is with the School of Electronics and Information Engineering, Guangdong University of Petrochemical Technology, Maoming 525000, China.

Shigang Yue is with the School of Computing and Mathematical Sciences, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: sy237@leicester.ac.uk).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2023.3345153, provided by the authors.

Digital Object Identifier 10.1109/TIP.2023.3345153

extracted. Traditionally, object motion detection is approached in motion-based methods or appearance-based methods. The former relies primarily on luminance changes of each pixel over time to estimate object motion and can be further subdivided into background subtraction [3], temporal differencing [4], and optical flow [5]. The later focuses on utilizing machine learning algorithms to learn visual features of objects in individual image and then match the extracted features in consecutive images to infer object movement [6]. However, these conventional methods would always suffer from two major problems when directly applied to the task of small target motion detection. First, distinctive visual feature representations are extremely difficult to be extracted from poor-quality appearances of small objects which are only one or a few pixels in size. In addition, background variations induced by mobile cameras highly degenerate their performance, since minute errors resulting form frame registration easily submerge small targets and lead to a number of false-positive detections. Consequently, effective solutions are required to bridge the performance gap in small target motion detection.

传统方法对小目标的不适用

Despite their low-resolution eyes and tiny brains, insects provide an elegant solution to discern small moving targets against complex dynamic environment robustly with limited computational resources [7]. For example, dragonflies are quite apt at chasing small mates or prey while performing sophisticated fast aerial maneuvers, evidenced by extremely high successful capture rate over 95% [8]. A group of neuron subtypes in insects' visual systems, called small target motion detectors (STMDs), are believed to underlie such excellent sensitivity [9], [10], [11]. They are strongly excited by movement of small targets subtending between 1° and 3° of the visual field, whereas moving objects subtending larger than 10° always elicit much weaker neural responses. Moreover, the STMD neurons respond robustly to small target motion even in complex dynamic environment, giving neither excitatory nor inhibitory responses to background clutters. Learning from insects' visual system and its neural implementation is clearly a promising way forward in developing robust and efficient image processing methods for small target motion detection.

To reproduce the superior properties of the STMD neurons in image processing, considerable efforts have been undertaken to establish explicit correspondence between insects' neural circuits and artificial visual systems. For example, an elementary STMD (ESTMD) model [12] was proposed to describe the remarkable selectivity of the STMD neurons for object sizes. To account for direction selectivity, cascaded models [13] and directionally selective STMD (DSTMD) [14] were successively developed. Combining the DSTMD with a contrast pathway in parallel, STMD plus [15] was designed to reveal the effect of information fusion on motion discrimination. These STMD-based models are all composed of multiple neural layers interconnected in a feed-forward way to process visual stimuli sequentially. Although these models can partly explain underlying neural computation of the STMD, their outputs often contain a number of background false positives from which real small target motion is difficult to be separated. The robustness of these models needs to be improved in dealing with small target detection against complex natural background. To suppress false positive background movement, time-delay feedback connection was introduced into the STMD-based methods [16]. However, **the time-delay feedback model** makes certain assumption that small targets should be faster than background, which means **it can only filter out those slow-moving background false positives**.

背景假阳性仍旧无法解决

为了抑制假阳性已经有努力：时间延迟反馈链接

但缺陷在于：小目标必须比背景移动的快

To solve the above problems, we propose a STMD-based visual system with spatio-temporal feedback mechanism (called Spatio-Temporal Feedback STMD). In insects' visual systems, movement patterns of surroundings are perceivable as spatial and temporal variations of brightness on the retina, which in turn shape dynamics of feedback signals [17], [18], [19], [20], [21]. To enhance object motion of interest while suppressing distracting signals from cluttered background, spatio-temporal feedback is desired but has not been deeply explored [22], [23], [24]. The proposed visual system is mainly composed of **two complementary subnetworks** called Lobula Plate Tangential Cell (LPTC) and Small Target Motion Detector (STMD), respectively, where the LPTC subnetwork is designed for inferring spatio-temporal dynamics of cluttered background while the STMD subnetwork is developed for capturing small target motion. We devise a **spatio-temporal feedback connection between the two subnetworks**, where the spatio-temporal information about background movement is integrated with the output of the STMD subnetwork to suppress background false positives in a recurrent manner.

Contributions of this paper can be summarized as follows:

- We add a new LPTC subnetwork to infer spatio-temporal dynamics of complex background by neural population coding, where parameter sensitivities of the LPTC subnetwork are also carefully studied.
- We design a spatio-temporal feedback loop that integrates the outputs of the STMD and LPTC subnetworks as feedback signal to suppress background false positives.
- We demonstrate the proposed model overcomes the limitation of the time-delay feedback model [16] and performs well without requirement of that small targets are faster than surrounding backgrounds.

We organize the rest of this paper as follows. Section II briefly overviews the related works. Section III introduces the details of the spatio-temporal feedback visual system. Section IV reports extensive experiment results together with qualitative studies on synthetic and real-world image sequences. Finally, Section V provides concluding remarks.

II. Related Work

A. Bio-Inspired Motion Detection

Inspiration from neuroscience is a promising approach for designing artificial visual systems with requirement of a high level of efficiency and robustness but limited in computational and memory budget [25], [26], [27], [28], [29], [30]. It has attracted a great deal of interests and become an emerging research area with a number of practical applications, such as visually guided flights or landing [31], [32], autonomous navigation [33], [34], and collision detection [35], [36]. Our work

is mainly related to two types of widely investigated motion-sensitive neurons, called lobula plate tangential cells (LPTCs) [37], [38] and small target motion detectors (STMDs), whose biological properties and computational models are briefly discussed.

LPTC neurons found in lobula plate of insects' visual systems, exhibit strong preference to object motion occupying large parts of the visual field (called wide-field motion). It is initially modelled by an array of elementary motion detectors (EMDs) [39], each of which perceives object motion in a small part of the visual field. Specifically, a single EMD model relies on multiplication of luminance-change signals from two neighboring pixels to generate positive outputs to object movement in its preferred direction, one of which have been time-delayed for signal alignment in temporal domain. Furthermore, two EMDs can be combined in a mirror-symmetric manner to discriminate responses to objects moving along preferred direction and null direction [40]. Biological research revealed that luminance increase and decrease signals are processed in parallel by medulla neural pathways and then recombined by the LPTC neurons for motion detection [41]. Taking this new finding into account, researchers proposed to split the input luminance-change signal into two parallel channels, which encode luminance increments and decrements, respectively. These two channels are then combined in possible pairs, giving rise to several variants of the EMD, such as Weighted Quadrant Detector [42], Two Quadrant Detector [43], and Four Quadrant Detector [43].

STMD neurons are highly sensitive to movement of small objects that occupy $1°-3°$ of the visual field. Inspired by their physiological characteristics, there have been many attempts to design the STMD-based neural networks. For example, the elementary STMD (ESTMD) [12] model was proposed to identify locations of small moving objects by implementing lateral inhibition and signal correlation mechanisms on intensity changes at each pixel. Specifically, it first separates the signal of intensity changes at each pixel into increase and decrease parts using half-wave rectification, which are further laterally inhibited to suppress large object motion. After that, the increase component is correlated with the time-delay decrease component via a multiplying unit to simulate the STMD responses. The ESTMD model was later improved by Wiederman and O'Carroll [13] to reproduce direction selectivity of the STMD neurons. They proposed to cascade the ESTMD with the EMD [39], resulting in two directionally selective models, called EMD-ESTMD and ESTMD-EMD. Wang et al. [14] provided an alternative called DSTMD to generate direction selectivity by correlating luminance-change signals from two different pixels. Based on insects' multi-information fusion schemes, the STMD plus model [15] was developed by parallelly integrating spatial contrast with motion information to eliminate background features that are highly similar to small targets. However, the aforementioned models are all characterized by a feedforward hierarchy to process visual stimuli via multiple sequentially arranged neural layers. Although these feedforward models exhibit parts of selectivities the same as the STMD neurons, they generally yield a number of false positives in dealing with small target detection against complex dynamic environment. To suppress false positive background movement, a time-delay feedback mechanism was incorporated into the STMD-based neural network [16]. However, it can only filter out slow-moving background motion and would be powerless against those false positives with high speed.

对ESTMD操作的理解，为什么能识别小目标

本部老师们的成果

B. Spatio-Temporal Feedback

Motion perception in insects' visual systems involves propagation and transformation of visual information across multiple hierarchically organized neural layers each of which functionally specializes for processing various aspects of object motion [44]. These neural layers interact closely in both feedforward and feedback directions, where feedforward conveys visual signals to higher layers while feedback passes high-level semantic information down to lower layers for modulating neural coding, removing distracting signals, and optimizing motion estimation [45].

When an insect performs high-speed maneuvers against complex dynamic environments, its self-motion relative to surroundings casts spatio-temporal variations of brightness on the retina, making small target motion discrimination a huge challenge [17], [18]. However, insects still demonstrate remarkable abilities to robustly discern and tracks moving objects of interest, such as predators, prey, and conspecifics, against cluttered backgrounds. Biological research indicates that motion patterns of surroundings would shape spatial and temporal dynamics of feedback signals which in turn suppress neural responses to distracting background objects while enhancing those to small moving targets [19], [20], [21]. Moreover, such spatio-temporal feedback is propagated from the lobula plate tangential cells to target-selective neurons for filtering out background motion [46], [47], [48].

Integrating spatial and temporal information is essential in both insects' visual system functions and computer vision applications. For example, Yuan et al. [49], [50], [51] considered spatio-temporal memory and context aware in object tracking models, which significantly improved tracking performance. Liu et al. fused spatial and temporal features for anomaly detection [52], transportation activity recognition [53], and assistive driving perception [54], achieving performance comparable to state-of-the-art methods. Feedback has been focus of studies on network structure and proved to be capable of significantly boosting network performances in various tasks, such as visual segmentation [22], saliency detection [55], and object recognition [56]. Despite its success in these applications, feedback with specific spatio-temporal dynamics has not been incorporated in the STMD-based neural networks and its functional significance remains unclear.

C. Infrared Small Target Detection

Infrared small target detection aims to detect objects of interest with significant temperature difference to surrounding backgrounds, such as rockets, jets, and ships, at far distance by infrared imaging technology [57]. In recent years, many promising models have been proposed, achieving significant improvement in infrared small target detection task. For

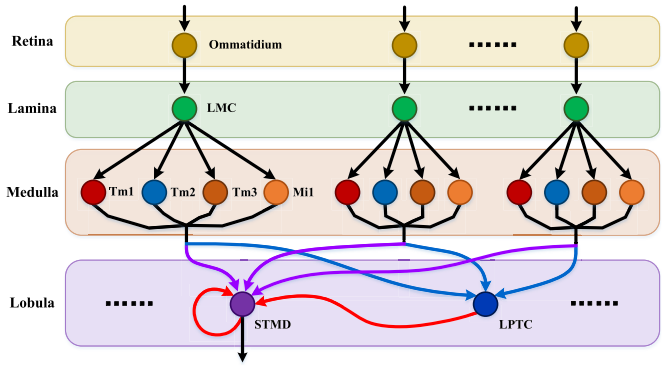


Fig. 2. Network structure of the proposed spatio-temporal feedback visual system. It is hierarchically organized and composed of four sequentially arranged neural layers, including retina, lamina, medulla, and lobula (from top to bottom). Each layer comprises a number of specialized neurons denoted by colored nodes. Red lines denote feedback loops while other colored lines represent feedforward connections.

example, Dai et al. [58] developed an asymmetric contextual modulation that integrates top-down and bottom-up attention to highlight features of infrared small targets. Li et al. [59] proposed a deny nested interactive module to maintain information of infrared small targets in deep layers. However, infrared-based methods are heavily dependent on temperature difference between objects and background, which means they would be unable to distinguish objects that irradiate similar thermal energy levels [60]. In addition, detection environment of the infrared-based methods is mainly sky or ocean, which are much more clear and homogeneous than the cluttered natural environments. In complex natural environment that contains a number of background features similar to small targets, such as such as bushes, trees, shadows, and light spots, the infrared-based methods may experience performance degradation and their outputs may contain many background false positives [61].

III. Spatio-Temporal Feedback Visual System

The proposed spatio-temporal feedback visual system is a multi-layer network with a number of specialized neurons, as shown in Fig. 2. External visual stimuli successively flow through ommatidia in retina layer [62], large monopolar cells (LMCs) in lamina layer [63], medulla neurons (e.g., Tm1, Tm2, Tm3, and Mi1) in medulla layer [64], and are integrated by the STMD [11] and LPTC [38] neurons in lobula layer, respectively. Besides the feedforward connections, the STMD neuron applies its output to its input in a recurrent manner, forming a self-feedback connection, while the LPTC neuron propagates its output incorporating spatio-temporal environmental dynamics to the input of the STMD neuron through an intra-layer feedback loop. We describe functionalities of each neural layer and their formulations in Sections III-A–III-D.

A. Retina Layer

As shown in Fig. 2, the retina layer consists of numerous optical units, ommatidia, each of which views a sector of the whole visual field and sends its axons to the lamina layer. The ommatidia act as luminance receptors with roughly Gaussian sensitivity profile to capture visible light from the natural scene [62]. The retina is designed as a two-dimensional array of ommatidia to receive the whole image frame, where each ommatidium is described as a Gaussian filter for yielding a smooth effect on pixel values, as can be seen from Fig. 3. Mathematically, given an image sequence $I(x, y, t) \in \mathbb{R}$ where x, y is spatial coordinate and t represents time, we formulate the output of an ommatidium $P(x, y, t)$ as

$$P(x, y, t) = \iint I(u, v, t) \cdot G_{\sigma_1}(x - u, y - v) du dv, \quad (1)$$

where $G_{\sigma_1}(x, y)$ denotes a Gaussian kernel with standard deviation σ_1, namely

$$G_{\sigma_1}(x, y) = \frac{1}{2\pi\sigma_1^2} \exp(-\frac{x^2 + y^2}{2\sigma_1^2}). \quad (2)$$

B. Lamina Layer

Large monopolar cells (LMCs) receive direct synaptic input from the ommatidia and then carry visual information to the medulla layer, as depicted in Fig. 2. The LMCs are highly sensitive to luminance changes over time, with hyperpolarizing and depolarizing responses to luminance increases and decreases, respectively [63]. Each LMC is modelled as a temporal filter to calculate luminance changes of each pixel in relation to time (see Fig. 3). The difference of two Gamma kernels [65] is adopted as the impulse response of the LMC, considering their excellent temporal processing properties, such as trivial stability, easy adaptation, and the uncoupling of impulse response and filter order. Let $H(t)$ denote the impulse response of the LMC, then we can write it as

$$H(t) = \Gamma_{n_1,\tau_1}(t) - \Gamma_{n_2,\tau_2}(t), \quad (3)$$

$$\Gamma_{n,\tau}(t) = (nt)^n \frac{\exp(-nt/\tau)}{(n-1)! \cdot \tau^{n+1}}, \quad (4)$$

where $\Gamma_{n,\tau}(t)$ stands for a Gamma kernel with order n and time constant τ. The output of the LMC $L(x, y, t)$ is defined by convolution of the ommatidium output $P(x, y, t)$ with $H(t)$

$$L(x, y, t) = \int P(x, y, s) \cdot H(t - s) ds. \quad (5)$$

Note that luminance of pixel (x, y) will change over time t when an object passes through it. Such temporal change in luminance is reflected in the output of the LMC $L(x, y, t)$. Specifically, the value of the LMC output $L(x, y, t)$ represents the amount of luminance change while the positive and negative signs correspond to luminance increase and decrease, respectively.

C. Medulla Layer

Each LMC innervates four medulla neurons, including Tm1, Tm2, Tm3, and Mi1, constituting four parallel information processing pathways [41], as can be seen from Fig. 2. Specifically, the Mi1 and Tm3 selectively respond to luminance increases with the Mi1 exhibiting a time-delayed response compared with the Tm3; the Tm1 and Tm2 selectively respond to luminance decreases with the Tm1 being temporally-delayed relative to the Tm2. In Fig. 3, the Tm3 and Tm2

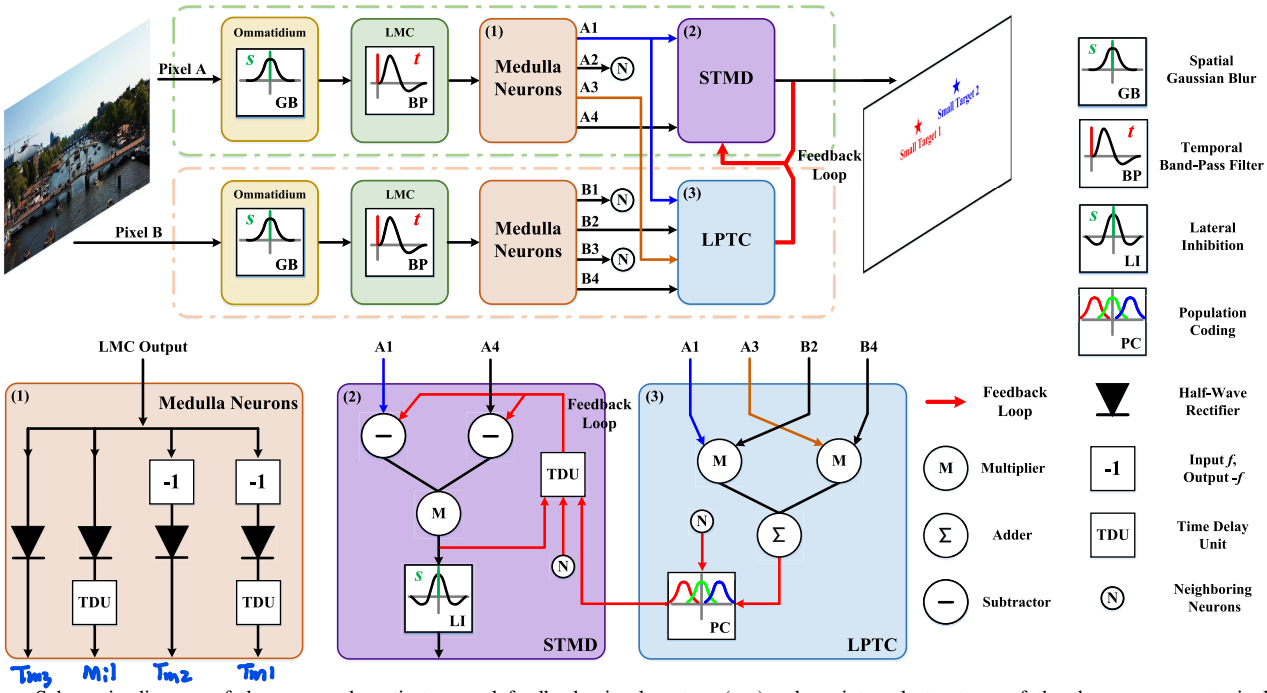


Fig. 3. Schematic diagram of the proposed spatio-temporal feedback visual system (top), where internal structures of the three components including medulla neurons, STMD, and LPTC, are illustrated at the bottom. Each type of specialized neurons in each layer is arranged in matrix form. To simplify the presentation, only one STMD, one LPTC, and their presynaptic neurons are shown. The entire image frame is applied to retina layer and further processed by the four neural layers in a feedforward manner. In lobula layer, the outputs of the STMD and LPTC are integrated together to act on the inputs of the STMD via feedback loops denoted by red lines. Specifically, spatio-temporal background dynamics are inferred by the LPTC neural population coding, further combined with the output of the STMD via a time-delay unit, finally propagated to the input of the STMD as feedback signal.

neurons are described as half-wave rectifiers to pass positive part and negative part of the LMC output $L(x, y, t)$, respectively. Denote the output of the Tm3 and Tm2 by $S^{\mathrm{Tm3}}(x, y, t)$ and $S^{\mathrm{Tm2}}(x, y, t)$, respectively, then they can be formulated as

$$S^{\mathrm{Tm3}}(x, y, t) = [L(x, y, t)]^+, \quad (6)$$

$$S^{\mathrm{Tm2}}(x, y, t) = [-L(x, y, t)]^+, \quad (7)$$

where $[x]^+$ represents $\max(x, 0)$. The output of the Mi1 and Tm1 neurons denoted by $S^{\mathrm{Mi1}}_{(n,\tau)}(x, y, t)$ and $S^{\mathrm{Tm1}}_{(n,\tau)}(x, y, t)$, respectively, are formulated as temporally delayed versions of $S^{\mathrm{Tm3}}(x, y, t)$ and $S^{\mathrm{Tm2}}(x, y, t)$, where time delay is achieved by convoluting with a Gamma kernel $\Gamma_{n,\tau}(t)$, as illustrated in Fig. 3. Formally, $S^{\mathrm{Mi1}}_{(n,\tau)}(x, y, t)$ and $S^{\mathrm{Tm1}}_{(n,\tau)}(x, y, t)$ are written as

$$S^{\mathrm{Mi1}}_{(n,\tau)}(x, y, t) = \int [L(x, y, s)]^+ \cdot \Gamma_{n,\tau}(t - s)ds, \quad (8)$$

$$S^{\mathrm{Tm1}}_{(n,\tau)}(x, y, t) = \int [-L(x, y, s)]^+ \cdot \Gamma_{n,\tau}(t - s)ds, \quad (9)$$

where time constant τ and order n of Gamma kernel $\Gamma_{n,\tau}(t)$ control length and order of the time delay unit, respectively.

D. Lobula Layer

As shown in Fig. 2, medulla neural outputs are integrated by two types of lobula neurons, including lobula plate tangential cells (LPTCs) and small target motion detectors (STMDs), which exhibit strong responses to wide-field motion and small target motion, respectively. Furthermore, the response of the STMD is mediated by intra-layer feedback from the LPTC and its self-feedback. In Fig. 3, the LPTC and STMD are formulated as two motion detectors to extract wide-field motion and small target motion by correlating luminance change patterns from medulla neurons. The intra-layer feedback and the self-feedback loops are designed to embed spatio-temporal background dynamics into the STMD output for eliminating false positive responses in a recurrent manner.

1) Small Target Motion Detector: Each STMD takes input from two medulla neurons located at a single pixel. Specifically, the outputs of the two medulla neurons, i.e., Tm1 and Tm3, are first mediated by subtracting a spatio-temporal feedback signal and further correlated together via a multiplier to generate significant outputs to small target movement, which is written as

$$D(x, y, t) = \left\{S^{\mathrm{Tm3}}(x, y, t) - F(x, y, t)\right\} \times \left\{S^{\mathrm{Tm1}}_{(n_3,\tau_3)}(x, y, t) - F(x, y, t)\right\}, \quad (10)$$

where $D(x, y, t)$ represents the correlation output of the STMD at pixel (x, y) and time t, while $F(x, y, t)$ denotes the spatio-temporal feedback signal formulated as

$$F(x, y, t) = \alpha \cdot \int_{\Omega} \Big\{D(x - \phi, y - \psi, t - s) + E(x - \phi, y - \psi, t - s)\Big\} \cdot \Gamma_{n_4,\tau_4}(s)ds, \quad (11)$$

where α is a feedback constant; $\phi(t, s)$ and $\psi(t, s)$ denote the x and y components of background dynamics propagated from the LPTC neurons, respectively, whose formulations are given in the following Sections; n_4 and τ_4 are the order and time constant of the Gamma kernel; $E(x, y, t)$ is the weighted

summation of the surrounding STMDs' outputs, whose weight function is given by

$$W_e(x, y) = \frac{1}{2\pi\eta^2} \exp(-\frac{x^2 + y^2}{2\eta^2}), \tag{12}$$

where constant η is controlled by the preferred target sizes of the central STMD, and $E(x, y, t)$ is expressed as

$$E(x, y, t) = \iint S^{\mathrm{Tm3}}(u, v, t) \cdot S^{\mathrm{Tm1}}(u, v, t) \cdot W_e(x - u, y - v) du dv. \tag{13}$$

The correlation output $D(x, y, t)$ is then convolved with a lateral inhibition kernel $W_s(x, y)$ for suppressing responses to large independently-moving objects, that is

$$Q(x, y, t) = \iint D(u, v, t) \cdot W_s(x - u, y - v) du dv, \tag{14}$$

where $Q(x, y, t)$ represents the STMD output after lateral inhibition and $W_s(x, y)$ is formulated as

$$W_s(x, y) = A \cdot [g(x, y)]^+ + B \cdot [g(x, y)]^-, \tag{15}$$

$$g(x, y) = G_{\sigma_2}(x, y) - e \cdot G_{\sigma_3}(x, y) - \rho, \tag{16}$$

where $[x]^+$ and $[x]^-$ represent $\max(x, 0)$ and $\min(x, 0)$, respectively, A, B, e and ρ are constant.

2) Motivation of Spatio-Temporal Feedback: Motion is generally more homogeneous within an object region, compared to other visual features, such as color, texture, and luminance [66]. Grouping motion clusters that share similar spatio-temporal characteristics in an image sequence would provide a powerful cue for object discrimination [67], [68], [69]. Moreover, point trajectories spanning multiple frames are always more robust to short-term variations compared to two-frame motion fields in the task of motion segmentation [70], [71]. Fig. 4 illustrates trajectories of a small target and its surrounding background in spatio-temporal space. The small target has relative movement to the complex background, so its trajectory is dissimilar to those of surrounding background, which could be clearly discriminated by observing x and y components of trajectories with respect to time t, as shown in Fig. 4(b) and (c). ***Motivated by this***, we design $\phi(t, s)$, $\psi(t, s)$ to reflect spatial and temporal dynamics of background motion, which correspond to x and y components of background trajectories, respectively. Substituting $\phi(t, s)$ and $\psi(t, s)$ into the spatio-temporal feedback signal $F(x, y, t)$ given by (11), the STMD responses to false positives with similar motion dynamics to complex background would be largely suppressed after the negative feedback.

It is worthy mentioning that the time-delay feedback model [16] only uses the time-delay version of the neural output as feedback signal, ignoring the spatio-temporal background dynamics $\phi(t, s)$ and $\psi(t, s)$. As shown in Fig. 6 and 7 of the reference [16], the time-delay feedback shows a preference for fast-moving objects. Specifically, after applying the time-delay feedback, the neural response to a slow moving object would be largely suppressed while that to a high-velocity object would be properly maintained. When background velocity is higher than that of small target, background false positives would receive much weaker suppression from

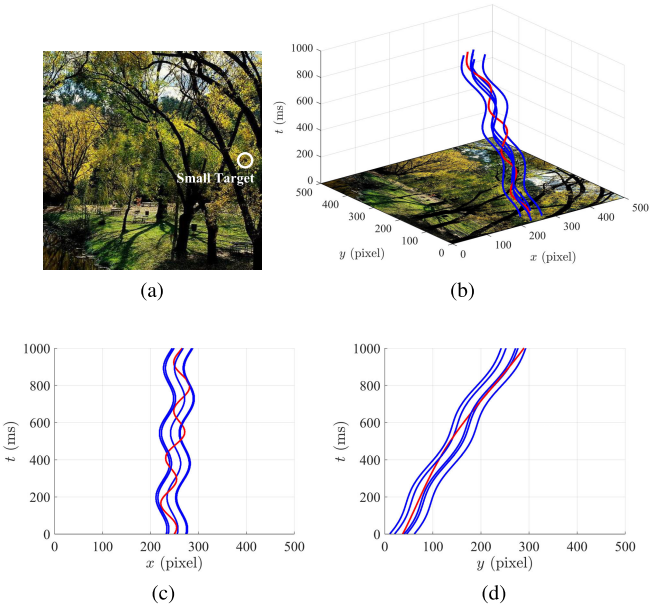


Fig. 4. (a) Representative image frame. (b) Object trajectories observed in spatio-temporal space. (c) Object trajectories observed in x-t plane. (d) Object trajectories observed in y-t plane. For the clarity of presentation, only one small target trajectory (red) and a few background object trajectories (blue) are shown. The red trajectory (small target) can be clearly discriminated from the blue trajectories (background objects) in terms of x and y components over time t, even though they seem to be extremely cluttered in spatio-temporal space.

the time-delay feedback loop, which consequently leads to performance degradation. In contrast, the spatio-temporal feedback model formulates feedback signal by accumulating the neural output based on background dynamics $\phi(t, s)$ and $\psi(t, s)$ (see Eq. (11)). After applying the spatio-temporal feedback, the neural responses with similar dynamics to the background would be significantly inhibited; on the contrary, the small target has relative movement to the background and its trajectory is dissimilar to those of the background (see Fig. 4), so the neural response to the small target would be maintained. That is, the spatio-temporal feedback improves detection performance by filtering out false positives with similar dynamics to background. In the following section, lobula plate tangential cell (LPTC) model and its population coding mechanism are proposed to obtain $\phi(t, s)$ and $\psi(t, s)$.

3) Lobula Plate Tangential Cell: Each LPTC collects outputs of medulla neurons located at two different pixels for wide-field motion detection, as shown in Fig. 3. Denote the two pixels by (x, y) and $(x'(\beta, \theta), y'(\beta, \theta))$, respectively, then they are written as

$$\begin{aligned} x'(\beta, \theta) &= x + \beta \cdot \cos\theta, \\ y'(\beta, \theta) &= y + \beta \cdot \sin\theta, \end{aligned} \tag{17}$$

where β is the distance between the two pixels and θ denotes the preferred direction of the LPTC. Multiplication of luminance-increase signals at these two pixels is summed together with that of luminance-decrease signals to define the output of the LPTC, formulated as

$$R(x, y, t, \beta, \theta) = S^{\mathrm{Tm3}}(x, y, t) \cdot S^{\mathrm{Mi1}}_{(n_5, \tau_5)}(x', y', t) + S^{\mathrm{Tm2}}(x, y, t) \cdot S^{\mathrm{Tm1}}_{(n_5, \tau_5)}(x', y', t), \tag{18}$$

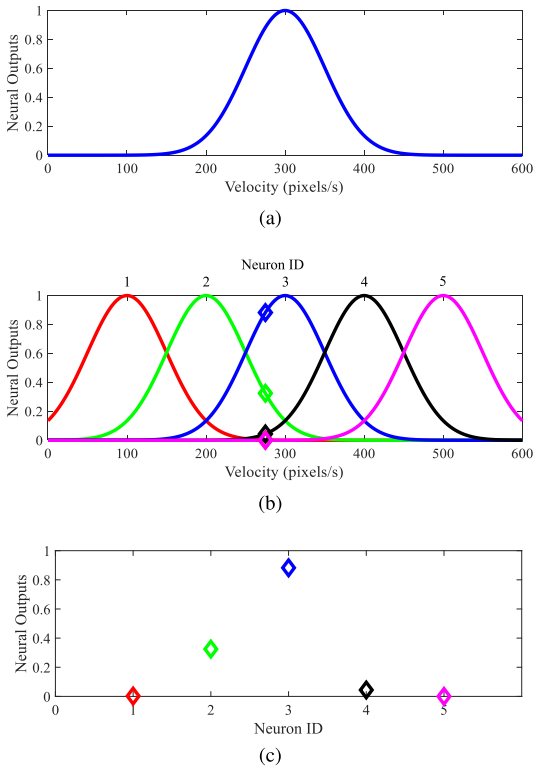


Fig. 5. (a) Tuning curve of a single neuron with respect to object velocity. The neuron responds to objects with velocities in a specific range roughly [150, 450] pixels/s, and its output peaks at an optimal velocity 300 pixels/s. (b) Schematic of neural population coding based on five neurons with different preferred velocity ranges. For an object with velocity 275 pixels/s, the intersection points with each tuning curve are arranged into a vector to yield the representation of object velocity as $\{r_k|k = 1, 2, \cdots, 5\} = \{0.0, 0.32, 0.88, 0.04, 0.0\}$. (c) Firing rate vector of all velocity tuned neurons $\{r_k|k = 1, 2, \cdots, 5\}$.

where $R(x, y, t, \beta, \theta)$ stands for the output of the LPTC; the time constants τ_5 is determined by time interval between the luminance-increase signals (or luminance-decrease signals) of the two pixels; the order n_5 controls the shapes of signals after the time delay.

Each LPTC is highly selective to object velocity. Specifically, it exhibits responses to the objects with velocities in a specific range and reaches the strongest output at an optimal velocity, as illustrated in Fig. 5(a). In animals' visual systems, object velocity is believed to be encoded by neural ensembles rather than an individual neuron [72], [73]. Taking inspiration from the neural population coding mechanism, we design a collection of the LPTC neurons with overlapping preferred velocity ranges to encode object velocity into neural activities, as shown in Fig. 5(b) and (c). More precisely, the output of the LPTC selective to a specific velocity range is obtained by adjusting the correlation distance between the two pixels, i.e., β in (18), whose sensitivity analysis is discussed in Section IV-B. We properly select a set of the correlation distance denoted by $\{\beta_i|i = 1, 2, \cdots, N\}$ to ensure the preferred velocity ranges of the LPTCs cover the object velocity range. Then the firing rate of the ith LPTC in response to a moving object is defined as the strongest output regarding to direction θ, that is

$$r(t, \beta_i) = \iint_{\Omega} R(x, y, t, \beta_i, \tilde{\theta}) dx dy, \tag{19}$$

$$\tilde{\theta} = \arg\max_{\theta} \iint_{\Omega} R(x, y, t, \beta_i, \theta) dx dy, \tag{20}$$

where $r(t, \beta_i)$ denotes the firing rate of the ith LPTC and $\tilde{\theta}$ represents the direction of the strongest output. The firing rate vector $\{r(t, \beta_i)|i = 1, 2, \cdots, N\}$ is further compared with the tuning curves of the LPTCs to estimate the velocity of a moving object, that is

$$v(t) = \arg\max_{v} \prod_{i} \exp\left(-||r(t, \beta_i) - f(v, \beta_i)||^2\right), \tag{21}$$

where $v(t)$ denotes the velocity of the moving object at time t and $f(v, \beta_i)$ represents the tuning curve of the ith LPTC. The x and y components of the object trajectory, i.e., $\phi(t, s)$ and $\psi(t, s)$, are defined using the accumulation of $v(t)$ during time period $[s, t]$, that is

$$\phi(t, s) = \int_{s}^{t} v(\tau) \cos\tilde{\theta} d\tau, \tag{22}$$

$$\psi(t, s) = \int_{s}^{t} v(\tau) \sin\tilde{\theta} d\tau. \tag{23}$$

The obtained $\phi(t, s)$ and $\psi(t, s)$ is further fed into the STMD subnetwork to incorporate in the spatio-temporal feedback signal $F(x, y, t)$.

IV. Experiments

A. Experimental Setup

1) Datasets: We quantify the performance of the developed spatio-temporal feedback visual system on Vision Egg dataset [74] and RIST dataset [75] for small target motion detection. Vision Egg is a large collection of video clips each of which holds one or multiple synthetic small targets moving against complex natural backgrounds. It contains a variety of background scene images and small targets categorized by luminance, velocity, and size. The spatial resolution of the videos range from 200 × 200 to 500 × 500 pixels, while their temporal resolutions are equal to 1000 fps. RIST is a challenging dataset consisting of 21 real-world video clips recorded by GoPro action camera with a spatial resolution of 480 × 270 pixels at 240 fps. Its scenarios involve many types of challenges, such as non-uniform dynamic shadows, heavy background clutters, sudden camera movement, various weather conditions, and illumination variations. The size of small targets in the captured videos ranges from 3 × 3 to 15 × 15 pixels. Average velocities of small targets in the two datasets all range from 0 to 1 pixel/frame.

2) Evaluation Metrics: We take detection rate (D_R) and false alarm rate (F_A) as metrics to measure the detection performance. Higher detection rate and lower false alarm rate denote better detection performance. The metrics can be formulated as follow

$$D_R = \frac{N_t}{N_a}, F_A = \frac{N_f}{N_F}, \tag{24}$$

where N_t is the number of true positives, N_a is the number of actual targets, N_f is the number of false positives, and N_F is the number of frames. A detected position is declared as a true positive if its distance to a ground truth is within 5 pixels. To assess detection performance under different detection

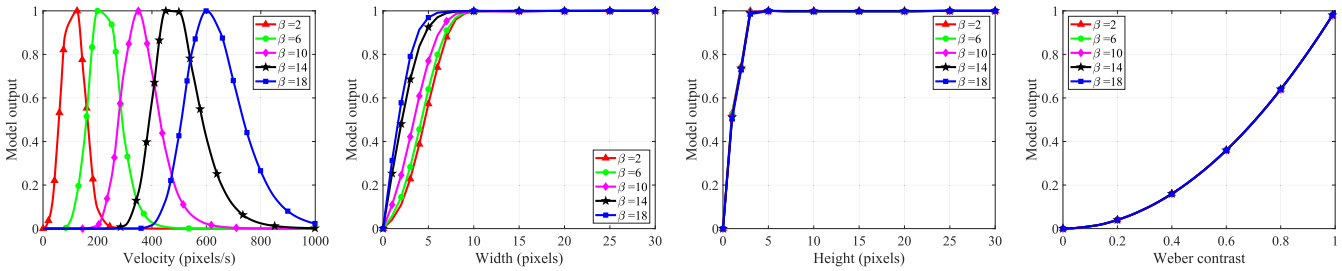


Fig. 6. Outputs of the LPTC neural model at various setting of the correlation distance β regarding to (a) velocity, (b) width, (c) height, (d) Weber contrast of the object.

TABLE I
PARAMETERS OF THE SPATIO-TEMPORAL FEEDBACK VISUAL SYSTEM

Eq.	Parameters
(1)	$\sigma_1 = 1$
(3)	$n_1 = 2, \tau_1 = 3, n_2 = 6, \tau_2 = 9$
(10)	$n_3 = 5, \tau_3 = 25$
(11)	$\alpha = 0.1, n_4 = 6, \tau_4 = 12$
(12)	$\eta = 1.5$
(15)	$A = 1, B = 3$
(16)	$\sigma_2 = 1.5, \sigma_3 = 3, e = 1, \rho = 0$
(17)	$\beta \in \{2, 4, 6, 8, 10, 12, 14, 16, 18\}$
(18)	$n_5 = 25, \tau_5 = 30$

thresholds, we adopt a receiver operating characteristic curve (ROC) which plots detection rate (D_R) against false positive rate (F_A) with a varying detection threshold.

3) Implementation Details: Parameters of the developed spatio-temporal feedback visual system are shown in Table I, which can be roughly classified into two groups for the STMD and LPTC subnetworks, respectively. The sensitivity analysis for the parameters of the STMD subnetwork and their effect on detection performance have been investigated in the previous works [14], [16]. In the experiments, we properly tune these parameters to endow the STMD subnetwork with similar response properties to the STMD neurons. As for the LPTC subnetwork, its performance is controlled by three parameters, including the correlation distance β, order n_5, and time constant τ_5 [see (17) and (18)]. Tuning these parameters will change the preferred velocity range of the LPTC. We further conduct a sensitivity study of these parameters in Section IV-B to evaluate their effect on preferred velocity range and provide a guidance for parameter setting of the population coding mechanism. The source code is publicly available at https://github.com/NIAIGroup/Spatio-Temporal-Feedback-STMD.

B. Parameter Sensitivity Study

Each LPTC neural model is determined by three parameters, including correlation distance β, order n_5, and time constant τ_5, as can be seen from (17) and (18). Their effects on the performance of the LPTC subnetwork are respectively evaluated in this section.

We first study the effect of the correlation distance β by tuning it in the range of $\{2, 6, 10, 14, 18\}$, while fixing $n_5 = 25$, $\tau_5 = 30$. Fig. 6(a)-(d) illustrates the outputs of the LPTC model under various setting of β regarding to (a) velocity, (b) width, (c) height, (d) Weber contrast of the object, respectively. Weber contrast measures the difference of average pixel intensity between an object and its surrounding region, whose formulation is given in the previous literature [16]. As shown in Fig. 6(a), for any given correlation distance β, the LPTC is highly selective to object velocity. Specifically, the LPTC exhibits positive responses in a specific velocity range (called preferred velocity range), and reach its maximal response at an optimal velocity. In addition, the velocity tuning curve changes significantly with respect to different settings of β. To be more precise, as β is increased from 2 to 18 pixels, the optimal velocity rises from 150 to 600 pixels/s while the preferred velocity range is also gradually shifted toward higher velocities. In Fig. 6(b) and (c), we can find that the LPTC does not exhibit strong selectivity for object width and height. For a fixed β, the response of the LPTC remains stable to object width larger than 10 pixels (or height larger than 5 pixels) after it reaches the maximum at width $=$ 10 pixels (or height $=$ 5 pixels). In addition, the responses with respect to object width and height are little affected by the increase of β, though the responses to width lower than 10 pixels shows a slight increase. From Fig. 6(d), we can see that the LPTC shows Weber contrast sensitivity where the increase in Weber contrast leads to the increase in the model output, which finally reaches its strongest response at Weber contrast $=$ 1. Furthermore, the Weber contrast sensitivity tuning curve remains unchanged at different settings of β.

We further study the performance variation of the LPTC model with respect to different values of the order n_5, where n_5 is tuned within $\{5, 10, 15, 20, 25\}$ while β and τ_5 are fixed to 4 and 30, respectively. As shown in Fig. 7(a), the decrease of the order n_5 results in the broader span of the preferred velocity while the optimal velocity remains almost unchanged. In Fig. 7(b), the increase of the order n_5 has small inhibitory effect on model outputs to object width lower than 10 pixels. However, once the width is greater than 10 pixels, changes in model output resulting from adjustment of n_5 will become little. From Fig. 7(c) and (d), we can observe that the order n_5 has minor influence on both height tuning curve and Weber contrast tuning curve.

Finally, we investigate the performance changes of the LPTC under different setting of the time constant τ_5 which is

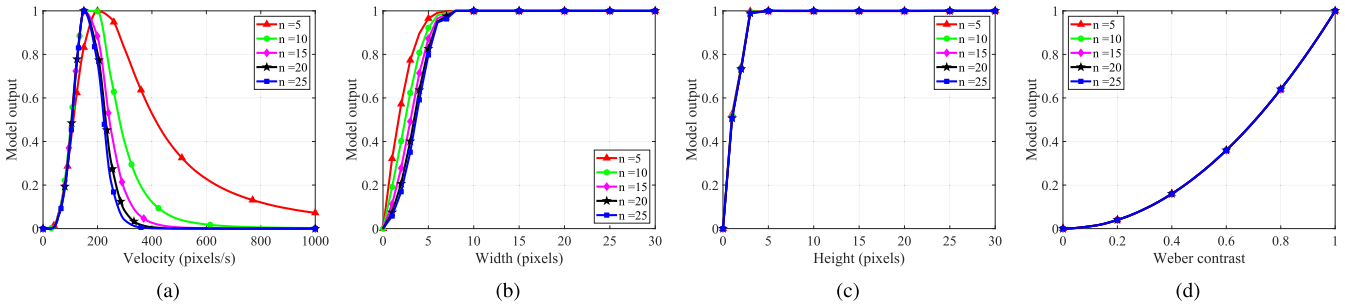


Fig. 7. Outputs of the LPTC neural model at various setting of the order n regarding to (a) velocity, (b) width, (c) height, (d) Weber contrast of the object.

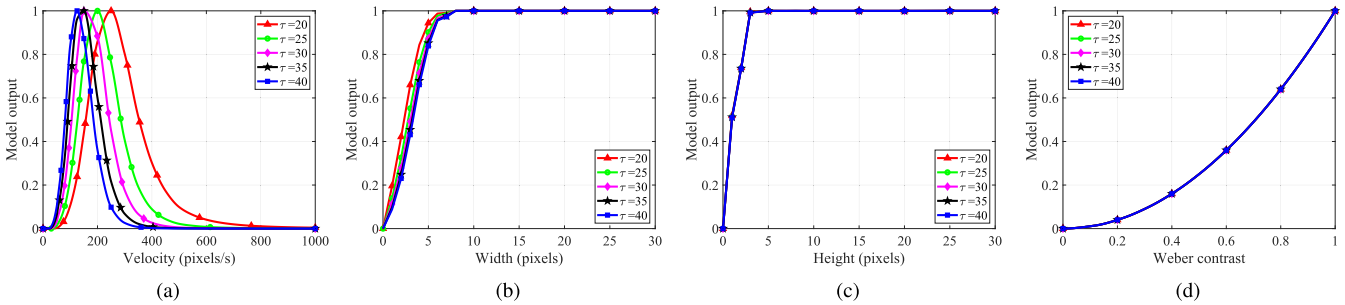


Fig. 8. Outputs of the LPTC neural model at various setting of the time constant τ regarding to (a) velocity, (b) width, (c) height, (d) Weber contrast of the object.

tuned within $\{20, 25, 30, 35, 40\}$ while β and n_5 are fixed to 4 and 15, respectively. From Fig. 8(a), it can be observed that the optimal velocity is slightly shifted toward higher values by the decrease in τ_5. The preferred velocity range is also broadened, where the maximum of the preferred velocity range increases to about 800 pixels/s when τ_5 decreases to 20, but the minimum remains unchanged at a low value about 50 pixels/s. As shown in Fig. 8(b)-(d), the three tuning curves are all insensitive to the changes of τ_5.

Based on the above sensitivity studies, we properly tune the correlation distance β to obtain multiple LPTCs with overlapped preferred velocity ranges for neural population coding. In Fig. 9, we further reveal the relationship between velocity estimated by the population coding mechanism and actual velocity under different setting of the LPTC neural number N. As can be seen, the increase in the number of the LPTCs will provide a more accurate estimation of object velocity within [150, 750] pixels/s. Compared to that of $N = 7$, the estimated velocity of $N = 9$ better fits the actual velocity around 400 pixels/s. The velocity estimation range of the population coding could be further broadened achieved by adding the LPTC neurons with preferred velocities higher than 750 pixels/s and/or lower than 150 pixels/s. In the experiments, we set the number of the LPTC neurons as 9 in the population coding mechanism for object velocity estimation.

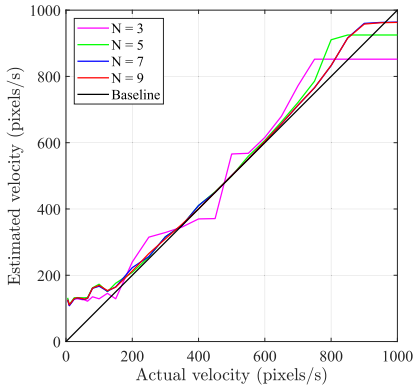


Fig. 9. Object velocity estimated by the population coding mechanism at various setting of neuron number N.

C. Effectiveness of the Neural Population Coding

We design a collection of the LPTC neurons with overlapping preferred velocity ranges to encode object velocity into neural activities. To validate its effectiveness, we conduct experiments on image sequences that hold multiple object motion with variable velocities. Fig. 10(a) illustrates velocities of two independently moving objects during time period [0, 1000] ms, where that of object 1 fluctuates frequently and significantly while that of object 2 exhibits a much more smooth change. The neural outputs of the LPTCs to these two objects over time are shown in Fig. 10(b) and (c), respectively. The outputs of the LPTCs are represented by colored rasters, where the redder the raster, the stronger response of the corresponding neuron. The larger neural ID means the corresponding LPTC neuron has a higher velocity range. As can be seen, a LPTC neuron will strongly fire only when object velocity is within its preferred velocity range; otherwise, it will fire sporadically or even not exhibit response. It can also be observed that the firing patterns of all the LPTC neurons for a moving object appears to encode the object velocity profile. Specifically, the ID of firing neurons are highly relevant to the object's velocity at a given time. Moreover, neural responses are quite reliable across the nine LPTCs for each moving object.

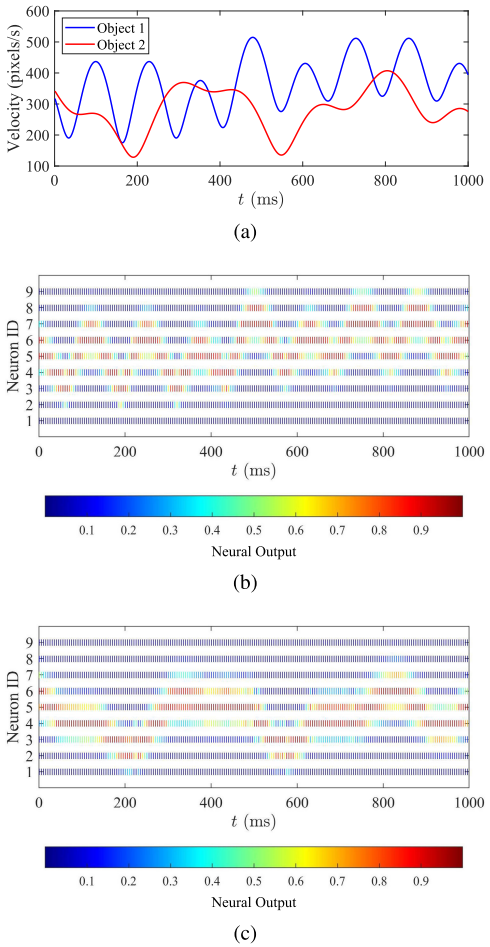


Fig. 10. (a) Velocities of two independently moving objects with respect to time t. (b) Raster plots of $N = 9$ LPTC neural outputs to the object 1. (c) Raster plots of $N = 9$ LPTC neural outputs to the object 2.

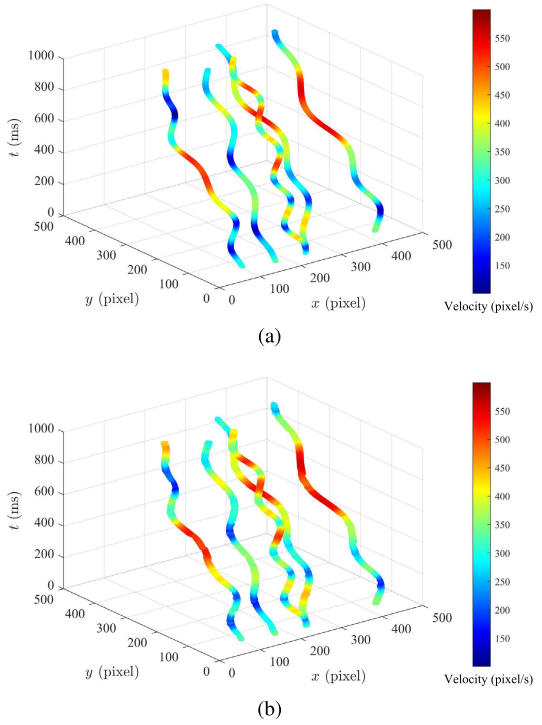


Fig. 11. (a) Actual trajectories and velocities, (b) trajectories and velocities estimated by the population coding mechanism in the three-dimensional (3D) view.

Fig. 12. Representative image that contains a small moving object without significant visual features while exhibiting extremely low contrast to complex environment. The size of the small object roughly equates to 5×5 pixels while its velocity approximates to 350 pixels/s which is lower than that of the surrounding background (450 pixels/s).

Fig. 11(a) shows actual trajectories and velocities of five independently moving objects where velocity is represented by color. For comparison purpose, trajectories and velocities estimated by the proposed population coding mechanism are shown in Fig. 11(b). We can observe that the population coding mechanism provides a good estimation for object velocities. Despite the trajectories having different velocities, accelerations, and angles, the estimation still can roughly match the ground truth given in Fig. 11(a). The above results demonstrate the effectiveness of neural population coding in velocity estimation, which plays a role in the LPTC subnetwork to extract spatio-temporal dynamics of background movement.

D. Effectiveness of the Spatio-Temporal Feedback

We develop a spatio-temporal feedback mechanism to suppress background false positives while enhancing responses to small moving targets. To validate its effectiveness, we conduct an ablation study by comparing the layer outputs of neural networks ***without feedback***, ***with time-delay feedback***, and ***with spatio-temporal feedback***. Fig. 12 shows a representative frame of the input video to the three neural networks at time $t = 500$ ms. In the scenario, a small target, lacking discriminative visual features and presenting extremely low contrast, is moving amidst clutters of natural environment. Its velocity roughly equates to 350 pixels/s whereas that of the surrounding background is about 450 pixels/s. For better observation of signal processing in each neural layer, we fix $y_0 = 125$ and present neural outputs in relation to x, where the input signal $I(x, y_0, t_0)$, the ommatidium output of the retina layer $P(x, y_0, t_0)$, and the LMC output of lamina layer $L(x, y_0, t_0)$, are displayed in Fig. 13(a)-(c), respectively. Because the time-delay feedback and spatio-temporal feedback are all implemented in the medulla and lobula layers, the three neural networks share the same ommatidium and LMC outputs. As shown, the ommatidium acts as a Gaussian blur achieving smoothing effects on the input signal while the LMC serves as a temporal filter to calculate luminance changes of each pixel over time. The positive and negative outputs of the LMC reflect luminance increase and decrease regarding to time, respectively.

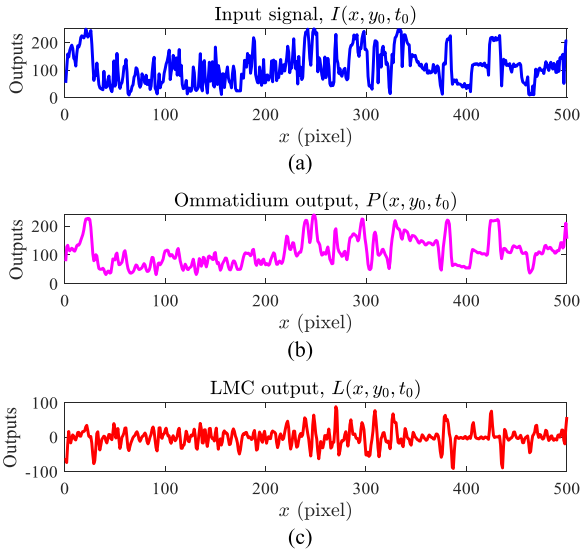


Fig. 13. (a) Input signal $I(x, y_0, t_0)$ with respect to x while fixing $y_0 = 125$ pixels and $t_0 = 500$ ms. (b) Ommatidium output of the retina layer $P(x, y_0, t_0)$. (c) LMC output of the lamina layer $L(x, y_0, t_0)$.

Fig. 14(a)-(c) presents medulla neural outputs without feedback, with time-delay feedback, and with spatio-temporal feedback, respectively. From Fig. 14(a), we can find that the outputs of two medulla neurons without feedback are derived from positive component and temporally-delayed negative component of the LMC output, respectively. Their product following by spatial lateral inhibition is used to define the output of the STMD neuron in Fig. 14(d). As can be seen, the STMD without feedback is highly responsive to the small target at $x = 128$, but it also exhibits significant outputs to background clutters, such as $x = 307$ and $x = 438$, due to mistaken correlation of luminance-change signals indued by background motion. To suppress these false-positive background responses, a feedback signal generated by introducing time lag into the STMD neural output is propagated to the lower layer for subtraction from the medulla neural outputs. The outputs of the medulla neurons to the small target and background clutters are all dramatically weakened after the time-delay feedback, as can be seen in Fig. 14(b), which eventually leads to noticeable decreases in the STMD neural outputs in Fig. 14(e). However, the time-delay feedback cannot completely eliminate background false positives with high velocities (see $x = 307$ and $x = 438$). To overcome the limitation of the time-delay feedback, the spatio-temporal feedback is designed by considering unique spatio-temporal characteristics of the background motion. As shown in Fig. 14(c), the two medulla neural outputs to background clutters are not only greatly suppressed but also properly aligned in time domain after the spatio-temporal feedback. These properly-aligned medulla neural outputs avoid mistaken correlation and finally generate the noiseless STMD neural output. As illustrated in Fig. 14(f), the STMD with the spatio-temporal feedback filter out all background false positives while slightly enhancing response to the small target. The above ablation study demonstrated the ability of the spatio-temporal feedback to suppress background false positives regardless of their velocities relative to small targets.

E. Comparative Results on Synthetic and Real Videos

We conduct quantitative evaluation on the synthetic and real-world datasets, i.e., Vision Egg [74] and RIST [75], in the matter of three key metrics, i.e., detection rate, false alarm rate, and ROC curve. Specifically, we first carry out experiments on five groups of synthetic videos each of which holds a single small target motion but differs in three object parameters (i.e., size, luminance, and velocity) and two background parameters (i.e., velocity and motion direction), to investigate their effects on detection performance. Secondly, we validate the effectiveness of our model for discriminating multiple small moving objects amidst various complex natural backgrounds. Finally, we perform the evaluation on the challenging real-world data set. Three state-of-the-art methods, including DSTMD [14], ESTMD [12], and Time-Delay Feedback STMD [16], are chosen for comparison, where their parameter settings are same with those in [12], [14], and [16].

1) Evaluations on Videos With Varying Parameters: The five parameters of the initial video, i.e., target size, target luminance, target velocity, background velocity, and background motion direction, are set as 5×5 pixels, 25, 250 pixels/s, 250 pixels/s, and leftward, respectively. Each group of synthetic videos tunes one of the parameters while fixing others at their initial values. Fig. 15(a) compares the ROC curves of the proposed spatio-temporal feedback model and baseline methods on the initial video. As shown, the spatio-temporal feedback model achieves the best performance. More precisely, it consistently has the highest detection rate close to 1 at a fixed false alarm rate, in comparison to other three baseline models.

To demonstrate superior performance of the spatio-temporal feedback under low false alarm rate, we further compare the detection rates of the models when $F_A = 5$, with respect to the five image parameters in Fig. 15(b)-(f), respectively. As illustrated in Fig. 15(b), the spatio-temporal feedback leads to significant performance improvements for all target sizes. In particular, its detection rate is close or even equal to 1 when target size ranges from 1×1 to 15×15 pixels, though experiences a sharp decrease for target size larger than 15×15 pixels. In contrast, the three competing models have a preferred size range roughly between 5×5 and 12×12 pixels where their maximal detection rates are much lower than that of the spatio-temporal feedback model. The result reveals that the spatio-temporal feedback is able to alleviate performance degradation induced by the size selectivity, but still cannot detect large objects with sizes larger than 20×20 pixels. From Fig. 15(c), we can see that the spatio-temporal feedback model achieves the best performance over the competing methods under various levels of luminance. Specifically, the spatio-temporal feedback maintains extremely high detection rate (close to 1) for target luminance ranging between 0 and 100. In contrast, the highest detection rates of the competing methods are all peaked at target luminance 0 and lower than 0.6. In addition, their detection performances substantially degrade with the increase in target luminance. The result indicates that the spatio-temporal feedback is able to alleviate the strong dependence on visual contrast between the small target and its surrounding background.

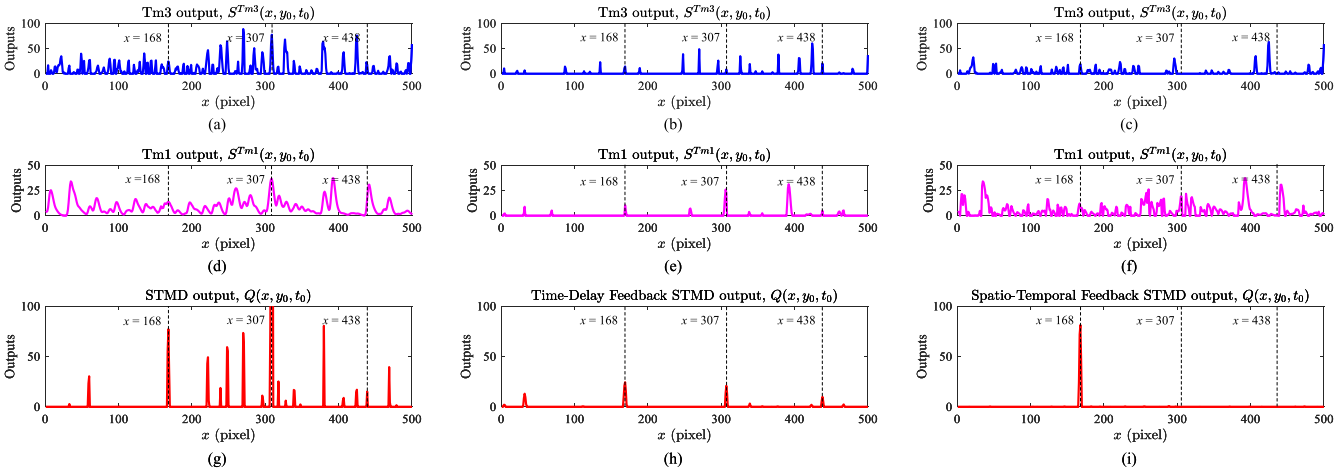


Fig. 14. (a) Output of the medulla neuron Tm3 without feedback. (b) Output of the medulla neuron Tm3 with time-delay feedback. (c) Output of the medulla neuron Tm3 with spatio-temporal feedback. (d) Output of the medulla neuron Tm1 without feedback. (e) Output of the medulla neuron Tm1 with time-delay feedback. (f) Output of the medulla neuron Tm1 with spatio-temporal feedback. (g) Output of the STMD neuron without feedback. (h) Output of the STMD neuron with time-delay feedback. (i) Output of the STMD neuron with spatio-temporal feedback.

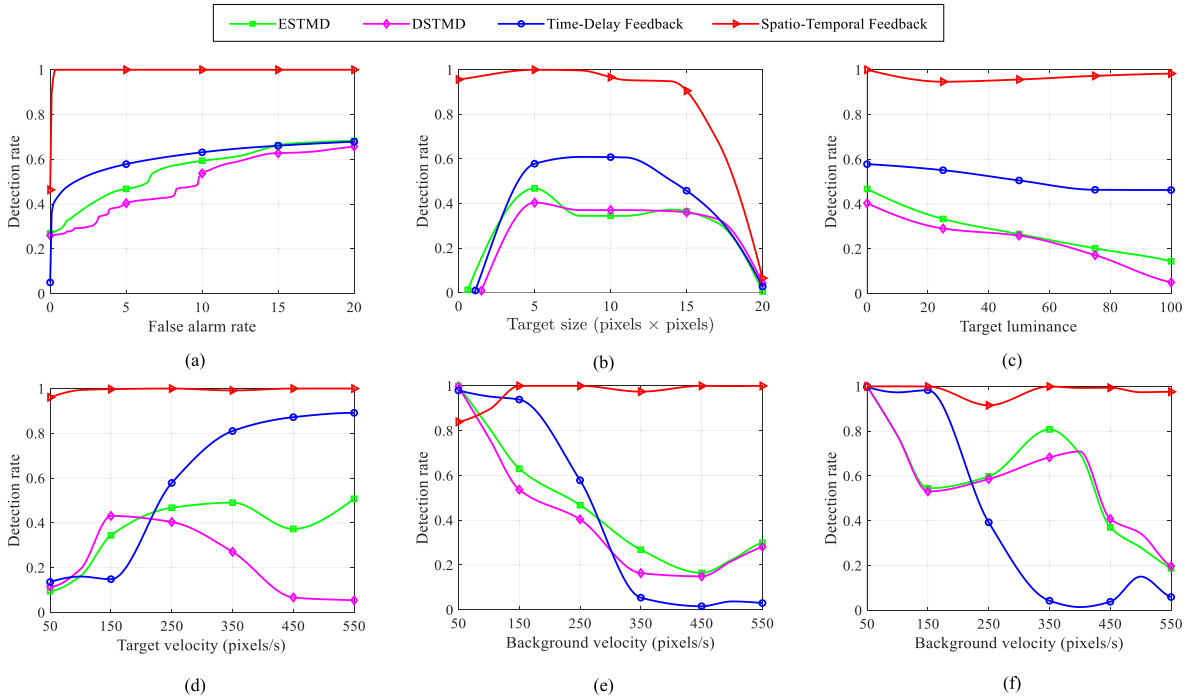


Fig. 15. (a) ROC curves of the four models on the initial video. (b)-(f) Detection rates of the four models when false alarm rate $F_A = 5$, in relation to (b) target size, (c) target luminance, (d) target velocity, (e) background velocity (leftward motion), and (f) background velocity (rightward motion).

As shown in Fig. 15(d)-(f), the spatio-temporal feedback model clearly outperforms the competing models under various target and background velocities. Its detection rate remains stable (close to 1) in most cases, while experiencing a slight decrease for background velocity lower than 150 pixels/s in leftward motion. Although the time-delay feedback improves the model performance in detecting those objects moving faster than the surroundings, it exhibits a significant performance degradation and performs the worst among the four models when the velocity of small target is lower than that of the background. *The above results demonstrate that the spatio-temporal feedback overcomes the limitation of the time-delay feedback and leads to significant performance improvements in all cases regardless of velocity difference between the small target and cluttered background.* Different background motion directions (leftward and rightward) would result in different relative motion between the small target and background, which consequently leads to difference in surrounding background regions, i.e., difference in contrast between the small target and background. The definition of contrast is consistent with that in [16] (see Eq.(19) and Fig. 11 of [16]). Due to different contrast of small target against background in leftward and rightward background motion, detection performance shows differences with respect to background direction.

2) Evaluations on Videos With Multiple Small Targets in Various Backgrounds: Fig.16(a)-(c) shows representative

(a) (b) (c)

Fig. 16. Representative images of the videos that hold (a) three, (b) four, and (c) five small targets moving in various complex backgrounds.

images of the synthetic videos with three, four, and five small moving targets, respectively. The ROC curve comparison are reported in *Supplementary Material*. The comparative results demonstrate that the spatio-temporal feedback achieves a substantial improvement over the other competing methods for detecting multiple small targets against various complex backgrounds. It has an obvious advantage over the competing models under any given false alarm rate.

3) Evaluations on Real-World Videos: We report the comparative results of the four methods in terms of the ROC curves on six real-world videos randomly selected from the RIST data set [75]. The comparative results are given in *Supplementary Material*, demonstrating that spatio-temporal feedback model achieves the best performance on all six real-world videos. Its detection rates are clearly higher than those of the competing methods by a large margin under any fixed false alarm rate.

F. Comparison With Infrared-Based Methods

We further compare the proposed model with two infrared-based methods including ACM [58] and DNANet [59] on a public infrared detection dataset [57]. The experimental results are given in *Supplementary Material*. As can be seen, all the three models show superior performance for detecting small targets against clean and homogeneous sky background. However, when coping with complex backgrounds that contain a number of features similar to small targets, DNANet and ACM experience significant performance degradation and produce many background false positives. Relying on the spatio-temporal feedback mechanism, the proposed model maintains significant responses to small moving targets while effectively suppressing background false positives.

V. Conclusion

In this paper, we proposed a biologically inspired image processing method for discriminating small target motion amidst heavy clutters of natural environment. The key difference from prior works lies in that a spatio-temporal feedback mechanism is devised to improve motion detection performance. Particularly, the spatio-temporal dynamics of background movement are integrated with the network output as a feedback signal, and then the elimination of background false positives is achieved by subtracting the feedback signal from the network input in a recurrent manner. We demonstrated that the proposed spatio-temporal feedback visual system obtained excellent performance against natural cluttered scenes. In future, other neural mechanisms, such as competition, attention, and evolution, could be coordinated together to further improve the robustness of small target motion detection in complex natural environment.

References

[1] *Amsterdam, Netherlands NL—By Drone*. Accessed: Sep. 5, 2020. [Online]. Available: https://www.youtube.com/watch?v=Oxv6IRcuNaI&list=FL_OHabe8rogCpinac5KHGYA&index=3&t=13s

[2] B. Webb, "Robots with insect brains," *Science*, vol. 368, no. 6488, pp. 244–245, Apr. 2020.

[3] B. Garcia-Garcia, T. Bouwmans, and A. J. R. Silva, "Background subtraction in real applications: Challenges, current models and future directions," *Comput. Sci. Rev.*, vol. 35, Feb. 2020, Art. no. 100204.

[4] L. Wang, Z. Tong, B. Ji, and G. Wu, "TDN: Temporal difference networks for efficient action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1895–1904.

[5] M. Zhai, X. Xiang, N. Lv, and X. Kong, "Optical flow and scene flow estimation: A survey," *Pattern Recognit.*, vol. 114, Jun. 2021, Art. no. 107861.

[6] T. Zhou, J. Li, S. Wang, R. Tao, and J. Shen, "MATNet: Motion-attentive transition network for zero-shot video object segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 8326–8338, 2020.

[7] M. S. Drews et al., "Dynamic signal compression for robust motion vision in flies," *Current Biol.*, vol. 30, no. 2, pp. 209–221, Jan. 2020.

[8] B. M. B. Bekkouche, P. A. Shoemaker, J. M. Fabian, E. Rigosi, S. D. Wiederman, and D. C. O'Carroll, "Modeling nonlinear dendritic processing of facilitation in a dragonfly target-tracking neuron," *Frontiers Neural Circuits*, vol. 15, Aug. 2021, Art. no. 684872.

[9] B. H. Lancer, B. J. E. Evans, J. M. Fabian, D. C. O'Carroll, and S. D. Wiederman, "A target-detecting visual neuron in the dragonfly locks on to selectively attended targets," *J. Neurosci.*, vol. 39, no. 43, pp. 8497–8509, Oct. 2019.

STMD 生理学

[10] K. Nordström, "Neural specializations for small target detection in insects," *Current Opinion Neurobiol.*, vol. 22, no. 2, pp. 272–278, Apr. 2012.

[11] J. M. Fabian and S. D. Wiederman, "Spike bursting in a dragonfly target-detecting neuron," *Sci. Rep.*, vol. 11, no. 1, pp. 1–6, Feb. 2021.

[12] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, "A model for the detection of moving targets in visual clutter inspired by insect physiology," *PLoS One*, vol. 3, no. 7, p. e2784, Jul. 2008.

ESTMD

[13] S. D. Wiederman and D. C. O'Carroll, "Biologically inspired feature detection using cascaded correlations of off and on channels," *J. Artif. Intell. Soft Comput. Res.*, vol. 3, no. 1, pp. 5–14, Jan. 2013.

带方向的 ESTMD-EMD

[14] H. Wang, J. Peng, and S. Yue, "A directionally selective small target motion detecting visual neural network in cluttered backgrounds," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1541–1555, Apr. 2020.

带方向的 DSTMD

[15] H. Wang, J. Peng, X. Zheng, and S. Yue, "A robust visual system for small target motion detection against cluttered moving backgrounds," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 839–853, Mar. 2020.

STMD+ 信息融合

[16] 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 Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 316–330, Jan. 2023.

[17] B. Cellini and J.-M. Mongeau, "Nested mechanosensory feedback actively damps visually guided head movements in drosophila," *eLife*, vol. 11, Oct. 2022, Art. no. e80880.

[18] O. W. Layton, E. Mingolla, and A. Yazdanbakhsh, "Neural dynamics of feedforward and feedback processing in figure-ground segregation," *Frontiers Psychol.*, vol. 5, p. 972, Sep. 2014.

[19] I. Uyanik et al., "Variability in locomotor dynamics reveals the critical role of feedback in task control," *eLife*, vol. 9, p. e51219, Jan. 2020.

[20] S. A. Stamper, E. Roth, N. J. Cowan, and E. S. Fortune, "Active sensing via movement shapes spatiotemporal patterns of sensory feedback," *J. Experim. Biol.*, vol. 215, no. 9, pp. 1567–1574, May 2012.

[21] S. C. Whitehead et al., "Neuromuscular embodiment of feedback control elements in *Drosophila* flight," *Sci. Adv.*, vol. 8, no. 50, Dec. 2022, Art. no. eabo7461.

[22] C. Cao, Y. Huang, Y. Yang, L. Wang, Z. Wang, and T. Tan, "Feedback convolutional neural network for visual localization and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1627–1640, Jul. 2019.

[23] H. Zhao et al., "Soft bimorph actuator with real-time multiplex motion perception," *Nano Energy*, vol. 76, Oct. 2020, Art. no. 104926.

[24] F. Paredes-Vallés, K. Y. W. Scheper, and G. C. H. E. de Croon, "Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2051–2064, Aug. 2020.

[25] G. C. H. E. de Croon, J. J. G. Dupeyroux, S. B. Fuller, and J. A. R. Marshall, "Insect-inspired AI for autonomous robots," *Sci. Robot.*, vol. 7, no. 67, Jun. 2022, Art. no. eabl6334.

[26] K.-F. Yang, X.-S. Zhang, and Y.-J. Li, "A biological vision inspired framework for image enhancement in poor visibility conditions," *IEEE Trans. Image Process.*, vol. 29, pp. 1493–1506, 2020.

[27] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171–177, Jan. 2010.

[28] I. Martins, P. Carvalho, L. Corte-Real, and J. L. Alba-Castro, "Bio-inspired boosting for moving objects segmentation," in *Proc. Int. Conf. Image Anal. Recognit. (ICIAR)*, 2016, pp. 397–406.

[29] A. Guzman-Pando and M. I. Chacon-Murguia, "DeepFoveaNet: Deep fovea eagle-eye bioinspired model to detect moving objects," *IEEE Trans. Image Process.*, vol. 30, pp. 7090–7100, 2021.

[30] M. Uzair, R. S. Brinkworth, and A. Finn, "Bio-inspired video enhancement for small moving target detection," *IEEE Trans. Image Process.*, vol. 30, pp. 1232–1244, 2021.

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

[32] M. Mischiati, H.-T. Lin, P. Herold, E. Imler, R. Olberg, and A. Leonardo, "Internal models direct dragonfly interception steering," *Nature*, vol. 517, no. 7534, pp. 333–338, Jan. 2015.

[33] S. Wang, Z. Qiu, P. Huang, X. Yu, J. Yang, and L. Guo, "A bioinspired navigation system for multirotor UAV by integrating polarization compass/magnetometer/INS/GNSS," *IEEE Trans. Ind. Electron.*, vol. 70, no. 8, pp. 8526–8536, Aug. 2023, doi: 10.1109/TIE.2022.3212421.

[34] X. Xiong and P. Manoonpong, "No need for landmarks: An embodied neural controller for robust insect-like navigation behaviors," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 12893–12904, Dec. 2022.

[35] 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 Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 5, pp. 2539–2553, May 2023, doi: 10.1109/TNNLS.2021.3106946.

[36] Y. Wang et al., "Memristor-based biomimetic compound eye for real-time collision detection," *Nature Commun.*, vol. 12, no. 1, p. 5979, Oct. 2021.

[37] J. C. Tuthill and B. G. Borghuis, "Four to foxtrot: How visual motion is computed in the fly brain," *Neuron*, vol. 89, no. 4, pp. 677–680, Feb. 2016.

[38] M. Henning, G. Ramos-Traslosheros, B. Gür, and M. Silies, "Populations of local direction-selective cells encode global motion patterns generated by self-motion," *Sci. Adv.*, vol. 8, no. 3, Jan. 2022, Art. no. eabi7112.

[39] M. Frye, "Elementary motion detectors," *Current Biol.*, vol. 25, no. 6, pp. 215–217, Mar. 2015.

[40] B. A. Badwan, M. S. Creamer, J. A. Zavatone-Veth, and D. A. Clark, "Dynamic nonlinearities enable direction opponency in *Drosophila* elementary motion detectors," *Nature Neurosci.*, vol. 22, no. 8, pp. 1318–1326, Jul. 2019.

[41] 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, Aug. 2014.

[42] D. A. Clark, L. Bursztyn, M. A. Horowitz, M. J. Schnitzer, and T. R. Clandinin, "Defining the computational structure of the motion detector in drosophila," *Neuron*, vol. 70, no. 6, pp. 1165–1177, Jun. 2011.

[43] H. Eichner, M. Joesch, B. Schnell, D. F. Reiff, and A. Borst, "Internal structure of the fly elementary motion detector," *Neuron*, vol. 70, no. 6, pp. 1155–1164, Jun. 2011.

[44] S. E. Kwon, H. Yang, G. Minamisawa, and D. H. O'Connor, "Sensory and decision-related activity propagate in a cortical feedback loop during touch perception," *Nature Neurosci.*, vol. 19, no. 9, pp. 1243–1249, Jul. 2016.

[45] S. E. Clarke and L. Maler, "Feedback synthesizes neural codes for motion," *Current Biol.*, vol. 27, no. 9, pp. 1356–1361, May 2017.

[46] S. Nicholas and K. Nordström, "Facilitation of neural responses to targets moving against optic flow," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 38, Sep. 2021, Art. no. e2024966118.

[47] C. Städele, M. F. Keleş, J.-M. Mongeau, and M. A. Frye, "Non-canonical receptive field properties and neuromodulation of feature-detecting neurons in flies," *Current Biol.*, vol. 30, no. 13, pp. 2508–2519, Jul. 2020.

[48] A. S. Mauss, K. Pankova, A. Arenz, A. Nern, G. M. Rubin, and A. Borst, "Neural circuit to integrate opposing motions in the visual field," *Cell*, vol. 162, no. 2, pp. 351–362, Jul. 2015.

[49] D. Yuan, X. Chang, Z. Li, and Z. He, "Learning adaptive spatial–temporal context-aware correlation filters for UAV tracking," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 18, no. 3, pp. 1–18, Aug. 2022.

[50] D. Yuan, X. Chang, P.-Y. Huang, Q. Liu, and Z. He, "Self-supervised deep correlation tracking," *IEEE Trans. Image Process.*, vol. 30, pp. 976–985, 2021.

[51] D. Yuan, X. Shu, Q. Liu, and Z. He, "Aligned spatial–temporal memory network for thermal infrared target tracking," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 70, no. 3, pp. 1224–1228, Mar. 2023.

[52] Y. Liu et al., "AMP-Net: Appearance-motion prototype network assisted automatic video anomaly detection system," *IEEE Trans. Ind. Informat.*, early access, Aug. 2023, doi: 10.1109/TII.2023.3298476.

[53] J. Liu, Y. Liu, W. Zhu, X. Zhu, and L. Song, "Distributional and spatial–temporal robust representation learning for transportation activity recognition," *Pattern Recognit.*, vol. 140, Aug. 2023, Art. no. 109568.

[54] D. Yang et al., "AIDE: A vision-driven multi-view, multi-modal, multi-tasking dataset for assistive driving perception," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 20459–20470.

[55] G. Li, Z. Liu, M. Chen, Z. Bai, W. Lin, and H. Ling, "Hierarchical alternate interaction network for RGB-D salient object detection," *IEEE Trans. Image Process.*, vol. 30, pp. 3528–3542, 2021.

[56] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, "Human action recognition from various data modalities: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3200–3225, Mar. 2023.

[57] S. Z. Hui-Bingwei, "A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background," *Sci. Data Bank*, vol. 5, p. 12, Sep. 2020.

[58] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 950–959.

[59] B. Li et al., "Dense nested attention network for infrared small target detection," *IEEE Trans. Image Process.*, vol. 32, pp. 1745–1758, 2023.

[60] S. Bianconi and H. Mohseni, "Recent advances in infrared imagers: Toward thermodynamic and quantum limits of photon sensitivity," *Rep. Prog. Phys.*, vol. 83, no. 4, Mar. 2020, Art. no. 044101.

[61] P. Yan, R. Hou, X. Duan, C. Yue, X. Wang, and X. Cao, "STD-MANet: Spatio-temporal differential multiscale attention network for small moving infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5602516.

[62] E. M. Caves, N. C. Brandley, and S. Johnsen, "Visual acuity and the evolution of signals," *Trends Ecol. Evol.*, vol. 33, no. 5, pp. 358–372, May 2018.

[63] 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," *Sci. Adv.*, vol. 6, no. 16, Apr. 2020, Art. no. eaaz8645.

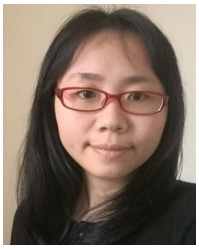
[64] K. Shinomiya et al., "Comparisons between the on-and off-edge motion pathways in the *Drosophila* brain," *eLife*, vol. 8, Jan. 2019, Art. no. e40025.
[65] B. De Vries and J. C. Príncipe, "A theory for neural networks with time delays," in *Proc. NIPS*, 1990, pp. 162–168.
[66] M. Keuper, S. Tang, B. Andres, T. Brox, and B. Schiele, "Motion segmentation & multiple object tracking by correlation co-clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 140–153, Jan. 2020.
[67] T. Stoffregen, G. Gallego, T. Drummond, L. Kleeman, and D. Scaramuzza, "Event-based motion segmentation by motion compensation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7244–7253.
[68] F. Arrigoni, E. Ricci, and T. Pajdla, "Multi-frame motion segmentation by combining two-frame results," *Int. J. Comput. Vis.*, vol. 130, no. 3, pp. 696–728, Jan. 2022.
[69] J. Bill, H. Pailian, S. J. Gershman, and J. Drugowitsch, "Hierarchical structure is employed by humans during visual motion perception," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 39, pp. 24581–24589, Sep. 2020.
[70] N. K. Kühn and T. Gollisch, "Activity correlations between direction-selective retinal ganglion cells synergistically enhance motion decoding from complex visual scenes," *Neuron*, vol. 101, no. 5, pp. 963–976, Mar. 2019.
[71] E. Levinkov, A. Kardoost, B. Andres, and M. Keuper, "Higher-order multicuts for geometric model fitting and motion segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 608–622, Jan. 2023.
[72] S. A. Koay, A. S. Charles, S. Y. Thiberge, C. D. Brody, and D. W. Tank, "Sequential and efficient neural-population coding of complex task information," *Neuron*, vol. 110, no. 2, pp. 328–349, Jan. 2022.
[73] J. I. Glaser, M. G. Perich, P. Ramkumar, L. E. Miller, and K. P. Kording, "Population coding of conditional probability distributions in dorsal premotor cortex," *Nature Commun.*, vol. 9, no. 1, pp. 1–14, May 2018.
[74] A. D. Straw, "Vision Egg: An open-source library for realtime visual stimulus generation," *Frontiers Neuroinform.*, vol. 2, no. 4, pp. 1–10, Nov. 2008.
[75] *RIST Data Set*. Accessed: Apr. 6, 2020. [Online]. Available: https://sites.google.com/view/hongxinwang-personalsite/download

Hongxin Wang received the B.Sc. and M.Sc. degrees in mathematics and applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 2013 and 2016, respectively, and the Ph.D. degree in computer science from the University of Lincoln, Lincoln, U.K., in 2020.

He was a Research Assistant with the Institute of Automation, Chinese Academy of Sciences, and Tsinghua University, Beijing, China, from 2015 to 2017, under the support of several EU FP7/Horizon2020 projects. He is currently a Lecturer with the Machine Life and Intelligence Research Center, Guangzhou University, Guangzhou, China. His current research interests include neuroscience-inspired artificial intelligence, biologically plausible information processing, and computational modeling.

Zhiyan Zhong received the Ph.D. degree in control theory and control engineering from the South China University of Technology, Guangzhou, China, in 2019.

She is currently an Associate Professor with the School of Automation, Guangdong Polytechnic Normal University, Guangzhou. Her research interests include industrial artificial intelligence and pattern recognition.



Fang Lei received the B.Sc. degree from the University of South China, Hengyang, China, in 2008, the M.Sc. degree from Hunan University, Changsha, China, in 2011, and the Ph.D. degree in computer science from the University of Lincoln, U.K., in 2022.

She is currently an Honorary Postdoctoral Research Fellow with the University of Lincoln and a Lecturer with the Guangdong University of Petrochemical Technology, Maoming, China. Her current research interests include biological/artificial vision systems and applications, computational neuroscience, and artificial intelligence.

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

He was a Lecturer with Xi'an Jiaotong University from 1992 to 1998, where he was an Associate Professor from 1998 to 2002 and a Professor from 2002 to 2018. He is currently a Professor with the Machine Life and Intelligence Research Center, School of Mathematics and Information Science, Guangzhou University, Guangzhou, China. His current research interests include non-linear functional analysis and applications, bio-plausible motion perception, and sparse information processing.

Shigang Yue (Senior Member, IEEE) received the B.Eng. degree from the Qingdao University of Technology, Qingdao, China, in 1988, and the M.Sc. and Ph.D. degrees from the Beijing University of Technology (BJUT), Beijing, China, in 1993 and 1996, respectively.

He was with BJUT as a Lecturer from 1996 to 1998 and an Associate Professor from 1998 to 1999. He was an Alexander von Humboldt Research Fellow with the University of Kaiserslautern, Kaiserslautern, Germany, from 2000 to 2001. He was a Senior Lecturer with the University of Lincoln, Lincoln, U.K., from 2007 to 2010, where he was a Professor of computer science from 2012 to 2023 and a Reader from 2010 to 2012. He held research positions with the University of Cambridge, Cambridge, U.K.; Newcastle University, Newcastle upon Tyne, U.K.; and University College London, London, U.K. He is currently a Professor of computer science with the School of Computing and Mathematical Sciences, University of Leicester, Leicester, U.K. He has been the Coordinator for several EU FP7/Horizon2020 projects. His current research interests include artificial intelligence, computer vision, robotics, brains and neuroscience, and their applications in autonomous vehicles, autonomous robots, interactive systems, and medical sciences.

Prof. Yue is a member of the International Neural Network Society and the International Symposium on Biomedical Engineering. He is a member of the Research Centre for Artificial Intelligence, Data Analytics and Modelling (AIDAM), Leicester. He was on the editorial board of several reputable journals in artificial intelligence, neural networks, and robotics.