

A Directionally Selective Small Target Motion Detecting Visual Neural Network in Cluttered Backgrounds

Hongxin Wang, Jigen Peng, Shigang Yue, *Senior Member, IEEE*

Abstract—Discriminating targets moving against a cluttered background is a huge challenge, let alone detecting a target as small as one or a few pixels and tracking it in flight. In the insect’s visual system, a class of specific neurons, called small target motion detectors (STMDs), have been identified as showing exquisite selectivity for small target motion. Some of the STMDs have also demonstrated direction selectivity which means these STMDs respond strongly only to their preferred motion direction. Direction selectivity is an important property of these STMD neurons which could contribute to tracking small targets such as mates in flight. However, little has been done on systematically modeling these directionally selective STMD neurons. In this paper, we propose a directionally selective STMD-based neural network for small target detection in a cluttered background. In the proposed neural network, a new correlation mechanism is introduced for direction selectivity via correlating signals relayed from two pixels. Then, a lateral inhibition mechanism is implemented on the spatial field for size selectivity of the STMD neurons. Finally, a population vector algorithm is used to encode motion direction of small targets. Extensive experiments showed that the proposed neural network not only is in accord with current biological findings, i.e., showing directional preferences, but also worked reliably in detecting small targets against cluttered backgrounds.

Index Terms—Cluttered backgrounds, direction selectivity, natural images, neural modeling, small target motion detection.

I. INTRODUCTION

INTELLIGENT robots have shown great potential in reshaping human life in the future. However, artificial visual systems so far are still struggling to provide a robot with the required capacity to respond to the dynamic visual world in real-time, like many animal species do. Among many visual functionalities, detecting small moving targets is one of the most important abilities for many animal species, e.g., finding

mates in the distance, and it is also critical for a robot to track small targets in a cluttered background.

Small target motion detection in visual cluttered backgrounds is always considered as a challenging problem for artificial visual systems. The difficulty is reflected in two aspects: first, when a target is far away from the observer, it always appears as a small dim speckle whose size may vary from one pixel to a few pixels in the field of view. In this size, shape, color and texture information cannot be used for target detection. Second, small targets are often buried in cluttered backgrounds and difficult to separate from noise. In addition, ego-motion may bring in further difficulties to small target motion detection.

Nature has provided a rich source of inspiration for small target motion detection. Detecting small targets in naturally cluttered backgrounds is critical for many insect species to search for mates or track prey. As the result of millions of years of evolution, the small target motion detection visual systems in insects are both efficient and reliable [1], [2]. For example, dragonflies can pursue small flying insects with successful capture rates as high as 97% relying on their well evolved vision system [3]. Compared to the visual systems of primate animals, insects’ visual systems achieve amazing capability using relatively simple structures and a small number of neurons. Insects’ visual pathways are practical models for designing artificial vision systems for small target motion detection.

In the insect’s visual system, a class of specific neurons, called small target motion detectors (STMDs), have been identified as showing exquisite selectivity for small targets (size selectivity) [2], [4], [5]. To be more precise, the STMD neurons give peak responses to targets subtending $1^\circ - 3^\circ$ of the visual region, with no response to larger bars (typically $> 10^\circ$) or to wide-field grating stimuli. In addition, some STMD neurons are directionally selective (direction selectivity) [6], [7]. They respond strongly to small target motion oriented along a preferred direction, but show weak or no, even fully opponent response to null-direction motion. Null direction is 180° from the preferred direction. Although the postsynaptic pathways of the STMD neurons are still under investigation [8], it is clear that knowing the small target motion and its direction at the same time is an advantage in tasks such as tracking mates or intercepting prey.

The electrophysiological knowledge about the STMD neurons and their afferent pathways revealed in the past few decades makes it possible to propose quantitative STMD

Manuscript received April 5, 2018; revised July 6, 2018 and July 18, 2018; accepted August 29, 2018. This work was supported in part by EU FP7 Project HAZCEPT under Grant 318907, in part by HORIZON 2020 project STEP2DYNA under Grant 691154, and in part by the National Natural Science Foundation of China under Grant 11771347. This paper is recommended by Associate Editor Q. Meng. (Hongxin Wang and Jigen Peng contributed equally to this work.) (*Corresponding author: Shigang Yue.*)

H. Wang and S. Yue are with the Computational Intelligence Lab (CIL), School of Computer Science, University of Lincoln, Lincoln LN6 7TS, U.K. (email: syue@lincoln.ac.uk).

J. Peng is with School of Mathematics and Information Science, Guangzhou University, Guangzhou, 510006, China (email: jgpeng@gzhu.edu.cn).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCYB.2018.2869384

models, however, little has been done on systematically modeling these directionally selective STMD neurons. As pioneers, Wiederman *et al.* [9] developed elementary small target motion detector (ESTMD) to account for size selectivity of the STMD neurons. The ESTMD showed strong responses to small target motion, but much weaker or even no responses to wide-field motion. However, it did not consider direction selectivity and showed no different responses to small target motion oriented along different directions. Wiederman and O'Carroll [10] mentioned that two hybrid models, i.e., elementary motion detector (EMD)-ESTMD and ESTMD-EMD, could exhibit both size and direction selectivities. In the further research [11]–[13], these two models are successfully used for target tracking. Although direction selectivity was noted in these models, the direction selectivity in an STMD model should be systematically investigated.

- 1) The existing STMD-based models, including ESTMD [9], EMD-ESTMD [10], and ESTMD-EMD [10], have not provided unified and rigorous mathematical description.
- 2) Wiederman and O'Carroll [10] and Bagheri *et al.* [11]–[13] focused on the size selectivity, tracking mechanisms and non-directionally selective properties, e.g., velocity and contrast tuning. Since direction selectivity has not been systematically studied, characteristics and performance of the directionally selective STMD models, are unclear.
- 3) The existing models have not shown the capacity for encoding motion direction of small targets.

In this paper, we propose a neural network to model the specific STMD neurons with direction selectivity called DSTMD. It can detect not only small target motion but also the motion direction in cluttered backgrounds. The proposed neural network incorporates a new correlation mechanism which correlates signals relayed from two pixels so as to introduce directional selectivity. Then, a lateral inhibition mechanism acting on correlation outputs is used for size selectivity. Finally, a population vector algorithm is used to encode motion direction of small targets. Systematic experiments are carried out to validate the proposed neural network in complex environments.

The main contributions of this paper can be summarized as follows.

- 1) We develop a new directionally selective STMD-based neural network (DSTMD) with unified and rigorous mathematical description.
- 2) We systematically study and test both directionally selective and non-directionally selective properties of the developed neural network.
- 3) We propose a population vector algorithm to encode motion direction of small targets.

The remainder of this paper is organized as follows. In Section II, the related work will be reviewed. In Section III, the proposed neural network is described in detail. In Section IV, the experiments are carried out to test the performances of the proposed neural network. We give further discussions in Section V and finally in Section VI, we conclude this paper.

II. RELATED WORK

In this section, we review the related work on three motion sensitive neurons, including the lobula giant movement detector (LGMD) [14]–[16], lobula plate tangential cell (LPTC) [17]–[19] and STMD [2], [4]–[7]. These three neurons are all found in insects' visual systems and have been extensively studied.

A. Lobula Giant Movement Detector

LGMDs are collision sensitive neurons found in locusts (certain species of short-horned grasshoppers) [14]–[16]. They respond strongly to the objects approaching the insect on a direct collision course while exhibiting little or no response to receding objects. A great number of LGMD-based neural networks [1], [20]–[24] have been developed. These neural networks showed the same collision sensitivity as the LGMD neuron and can detect collisions cheaply and reliably in a complex background. Nevertheless, they are incapable of detecting small target motion, and do not show size and direction selectivities.

B. Lobula Plate Tangential Cell

LPTCs exhibit strong responses to wide-field motion, but also to the motion of local, salient features [17]–[19]. The first LPTC model which is the spatial integration of elementary motion detectors (EMDs), was originally inferred from behavioral investigation of insects [25]. In the past decade, considerable progress has been made in identifying the afferent pathways and the characteristics of the LPTCs. To incorporate these new biological findings, the EMD was adapted, giving rise to several models, such as two-quadrant-detector (TQD) [26], [27] and weighted-quadrant-detector [28]. The above-mentioned models respond to objects' motion, but they are not size selective.

C. Small Target Motion Detector

Small target motion detectors are characterized by exquisite size selectivity [2], [4], [5], some of which are also directionally selective [6], [7]. Wiederman *et al.* [9] proposed the ESTMD, to model an STMD neuron with spatially integrated multiple ESTMDs. Although the ESTMD shows size selectivity, it is not directionally selective. Wiederman and O'Carroll [10] and Bagheri *et al.* [11]–[13] mentioned that two hybrid models, i.e., EMD-ESTMD and ESTMD-EMD, could exhibit both size and direction selectivities. However, characteristics and performance of these directionally selective STMD models, are unclear, since direction selectivity has not been systematically studied.

III. FORMULATION OF THE MODEL

Following the typical multi-stage view of motion detection in the insect's visual system (schematically illustrated in Fig. 1), we devised a DSTMD in this paper. Fig. 2(a) shows the schematic of one DSTMD cell and its presynaptic neural network. The proposed neural network is composed of four neural layers including the retina, lamina, medulla, and lobula. These

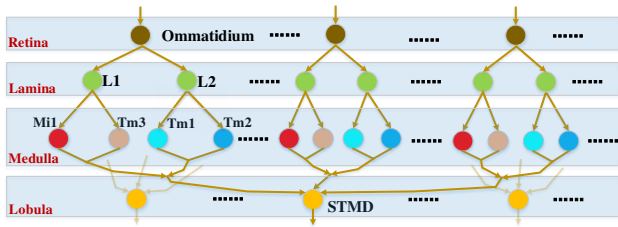


Fig. 1. Schematic illustration of the insect's visual system. The insect's visual system consists of four neural layers, including retina, lamina, medulla and lobula (from top to bottom). Each neural layer contains numerous specialized neurons illustrated by colored circular nodes. Luminance signals are first perceived by ommatidia, further processed by LMCs (i.e., L1 and L2) and medulla neurons (Mi1, Tm1, Tm2, Tm3), finally integrated in STMD neurons. Note that the connection between the four medulla neurons and the STMD neuron is speculative.

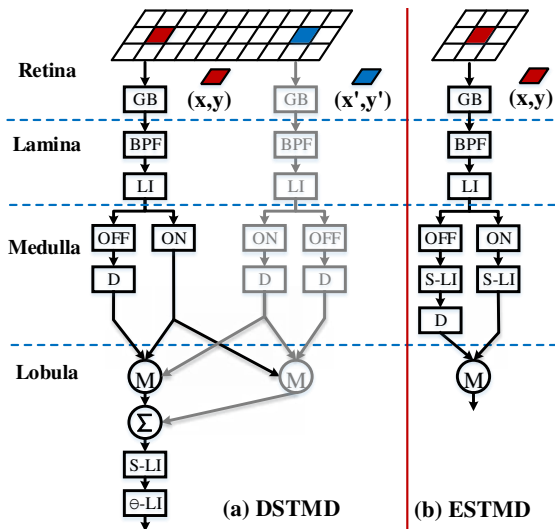


Fig. 2. Schematic illustration of the proposed DSTMD and the existing ESTMD models which exhibit selectivity for dark small targets. (a) Schematic illustration of one DSTMD located at (x, y) with a preferred direction θ . (b) Schematic illustration of one ESTMD located at (x, y) . The most significant difference between the DSTMD and ESTMD is that the DSTMD integrates signals from two different positions (x, y) and (x', y') where $x' = x + \alpha_1 \cos \theta$, $y' = y + \alpha_1 \sin \theta$, α_1 is a constant. However, the ESTMD integrates signals from a single position (x, y) . Therefore, for each position (x, y) , the DSTMD has multiple model outputs corresponding to different preferred directions θ while the ESTMD just has a single output without direction selectivity. *Abbreviation*, GB: Gaussian blur, BPF: band-pass filter, LI: lateral inhibition, ON/OFF: ON/OFF signals, D: time delay, M and Σ : multiplier and adder, S-LI: second-order lateral inhibition, θ -LI: lateral inhibition implemented on θ .

four sequentially arranged neural layers have specific functions and cooperate together for small target motion detection. In the following sections, we will elaborate on the components and functions of each layer.

A. Retina Layer

In the insect's visual system, the retina layer contains a great number of ommatidia (see Fig. 1). Each ommatidium is composed of eight photoreceptors. Each photoreceptor views a small region of the whole visual field and supplies a 'pixel' of luminance information to ommatidia [29].

In the proposed neural network, image sequences are network inputs, so we first construct a mapping from pixels

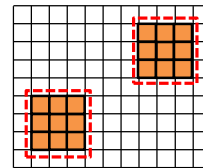


Fig. 3. Schematic illustration of the mapping from pixels to photoreceptors. Each small square denotes a pixel, corresponding to a photoreceptor. Each red dotted rectangle which contains multiple pixels (photoreceptors), represents the visual region of an ommatidium.

to photoreceptors. As depicted in Fig. 3, each small square denotes a pixel, corresponding to a photoreceptor. The red dotted rectangle which contains multiple pixels (photoreceptors), represents the visual region of an ommatidium.

Specifically, let $I(x, y, t) \in \mathbb{R}$ denote varying luminance values captured by photoreceptors where x, y and t are spatial and temporal field positions. Then the response of an ommatidium is approximated by Gaussian blur. That is, the output of an ommatidium with visual region centered at (x, y) denoted by $P(x, y, t)$ is given by,

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

where $G_{\sigma_1}(x, y)$ is a Gaussian function, defined as

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

B. Lamina Layer

In the insect's visual system, large monopolar cells (LMCs), such as L1 and L2, are postsynaptic neurons of the ommatidia (see Fig. 1). They receive signals from the ommatidia and show strong responses to luminance increments and decrements, i.e., luminance changes [30], [31].

In the proposed neural network, each LMC is modeled as a temporal band-pass filter to extract luminance changes from input signals. Let $L(x, y, t)$ denote the output of a LMC located at (x, y) . Then $L(x, y, t)$ is defined by convolving the ommatidium output $P(x, y, t)$ with a temporal band-pass filter $H(t)$. That is,

$$L(x, y, t) = \int P(x, y, s) H(t - s) ds \quad (3)$$

$$H(t) = \Gamma_{n_1, \tau_1}(t) - \Gamma_{n_2, \tau_2}(t) \quad (4)$$

where $\Gamma_{n, \tau}(t)$ is a Gamma kernel, defined as

$$\Gamma_{n, \tau}(t) = (nt)^n \frac{\exp(-nt/\tau)}{(n-1)! \tau^{n+1}}. \quad (5)$$

The illustration of the Gamma kernel $\Gamma_{n, \tau}(t)$ and temporal band-pass filter $H(t)$ is presented in Fig. 4.

In the insect's visual system, the LMC receives lateral inhibition from adjacent neurons before relaying its output to the next layer. In the proposed neural network, $L(x, y, t)$ is convolved with an inhibition kernel $W_1(x, y, t)$ so as to implement the lateral inhibition mechanism. That is,

$$L_I(x, y, t) = \iiint L(u, v, s) W_1(x - u, y - v, t - s) du dv ds \quad (6)$$

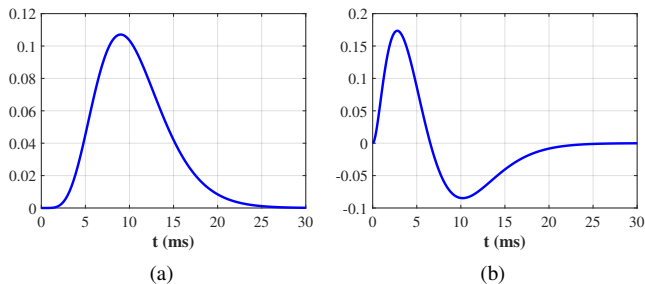


Fig. 4. (a) Gamma kernel $\Gamma_{n,\tau}(t)$ where $n = 6$, $\tau = 9$. (b) Temporal band-pass filter $H(t)$ where $n_1 = 2$, $\tau_1 = 3$, $n_2 = 6$, $\tau_2 = 9$.

where $L_I(x, y, t)$ is the signal after lateral inhibition and $W_1(x, y, t)$ is defined as,

$$W_1(x, y, t) = W_S^P(x, y)W_T^P(t) + W_S^N(x, y)W_T^N(t) \quad (7)$$

where $W_S^P(x, y)$, $W_S^N(x, y)$, $W_T^P(t)$, $W_T^N(t)$ are set as

$$W_S^P = [G_{\sigma_2}(x, y) - G_{\sigma_3}(x, y)]^+ \quad (8)$$

$$W_S^N = [G_{\sigma_2}(x, y) - G_{\sigma_3}(x, y)]^-, \quad \sigma_3 = 2 \cdot \sigma_2 \quad (9)$$

$$W_T^P = \frac{1}{\lambda_1} \exp\left(-\frac{t}{\lambda_1}\right) \quad (10)$$

$$W_T^N = \frac{1}{\lambda_2} \exp\left(-\frac{t}{\lambda_2}\right), \quad \lambda_2 > \lambda_1. \quad (11)$$

Here, $[x]^+$, $[x]^-$ denote $\max(x, 0)$ and $\min(x, 0)$, respectively.

C. Medulla Layer

In the insect's visual system, medulla neurons including Tm1, Tm2, Tm3 and Mi1, are downstream neurons of the LMCs (see Fig. 1). The Mi1 and Tm3 respond selectively to luminance increments, with the response of the Mi1 delayed relative to the Tm3 [31]. Conversely, the Tm1 and Tm2 respond selectively to luminance decrements, with the response of the Tm1 delayed relative to the Tm2 [32].

In the proposed DSTMD and the existing ESTMD [9], the modeling methods for these four medulla neurons, are different. These two modeling methods are introduced as follows, respectively.

1) *Medulla Neuron Modeling of DSTMD*: As the Tm3 and Tm2 neurons respond strongly to luminance increments and decrements, we use the positive and negative parts of the LMC output $L_I(x, y, t)$ to define the outputs of the Tm3 and Tm2, denoted by S^{Tm3} and S^{Tm2} , respectively. That is,

$$S^{\text{Tm3}}(x, y, t) = S^{\text{ON}}(x, y, t) \quad (12)$$

$$S^{\text{Tm2}}(x, y, t) = S^{\text{OFF}}(x, y, t) \quad (13)$$

where S^{ON} and S^{OFF} represent the positive and negative parts of $L_I(x, y, t)$, respectively. That is,

$$S^{\text{ON}}(x, y, t) = [L_I(x, y, t)]^+ \quad (14)$$

$$S^{\text{OFF}}(x, y, t) = -[L_I(x, y, t)]^- \quad (15)$$

where $S^{\text{ON}}(x, y, t)$ and $S^{\text{OFF}}(x, y, t)$ are also called ON and OFF signals (see the ON and OFF in Fig. 2), which reflect luminance increase and decrease, respectively.

Since the Mi1 (or Tm1) is a temporally delayed version of the Tm3 (or Tm2), the output of the Mi1 (or Tm1) is defined by convolving $S^{\text{Tm3}}(x, y, t)$ (or $S^{\text{Tm2}}(x, y, t)$) with a Gamma kernel. That is,

$$S_{D(n_N, \tau_N)}^{\text{Mi1}}(x, y, t) = \int S^{\text{Tm3}}(x, y, s) \Gamma_{n_N, \tau_N}(t - s) ds \quad (16)$$

$$S_{D(n_F, \tau_F)}^{\text{Tm1}}(x, y, t) = \int S^{\text{Tm2}}(x, y, s) \Gamma_{n_F, \tau_F}(t - s) ds \quad (17)$$

where $S_{D(n_N, \tau_N)}^{\text{Mi1}}$ and $S_{D(n_F, \tau_F)}^{\text{Tm1}}$ represent the outputs of the Mi1 and Tm1, respectively. n_N, n_F are orders of the Gamma kernels while τ_N, τ_F are time constants.

2) *Medulla Neuron Modeling of ESTMD*: The most significant difference between the medulla neuron modeling methods of the DSTMD and ESTMD is that the ESTMD uses laterally inhibited ON and OFF signals to define the outputs of the medulla neurons. This can be seen in Fig. 2 that the ESTMD implements a second-order lateral inhibition mechanism following ON and OFF signals while the DSTMD does not. In the ESTMD, the outputs of the Tm3 and Tm2 denoted by \tilde{S}^{Tm3} and \tilde{S}^{Tm2} , are defined as,

$$\tilde{S}^{\text{Tm3}}(x, y, t) = \left[\iint S^{\text{ON}}(u, v, t) W_2(x - u, y - v) dudv \right]^+ \quad (18)$$

$$\tilde{S}^{\text{Tm2}}(x, y, t) = \left[\iint S^{\text{OFF}}(u, v, t) W_2(x - u, y - v) dudv \right]^+ \quad (19)$$

where S^{ON} and S^{OFF} are the ON and OFF signals defined in (14) and (15); $W_2(x, y)$ is the second-order lateral inhibition kernel, defined as

$$W_2(x, y) = A[g(x, y)]^+ + B[g(x, y)]^- \quad (20)$$

where A, B are constant, and $g(x, y)$ is given by

$$g(x, y) = G_{\sigma_4}(x, y) - e \cdot G_{\sigma_5}(x, y) - \rho \quad (21)$$

where $G_{\sigma}(x, y)$ is a Gaussian function and e, ρ are constant.

Similarly, the outputs of the Tm1 and Mi1 are defined as the temporally delayed outputs of the Tm3 and Tm2, which are obtained by convolving $\tilde{S}^{\text{Tm3}}(x, y, t)$ and $\tilde{S}^{\text{Tm2}}(x, y, t)$ with a Gamma kernel. That is,

$$\tilde{S}_{D(n_N, \tau_N)}^{\text{Mi1}}(x, y, t) = \int \tilde{S}^{\text{Tm3}}(x, y, s) \Gamma_{n_N, \tau_N}(t - s) ds \quad (22)$$

$$\tilde{S}_{D(n_F, \tau_F)}^{\text{Tm1}}(x, y, t) = \int \tilde{S}^{\text{Tm2}}(x, y, s) \Gamma_{n_F, \tau_F}(t - s) ds \quad (23)$$

where $\tilde{S}_{D(n_N, \tau_N)}^{\text{Mi1}}$ and $\tilde{S}_{D(n_F, \tau_F)}^{\text{Tm1}}$ stand for the outputs of the Mi1 and Tm1, respectively.

In the following, we discuss the implementation of the second-order lateral inhibition mechanism. Existing biological research [33] asserts that the size selectivity of the STMD neurons is shaped by the second-order lateral inhibition mechanism. However, where this second-order lateral inhibition mechanism occurs remains elusive. Although the ESTMD implements this second-order lateral inhibition mechanism on medulla neurons, it is just speculative and there is no neuroanatomical evidence for it. On the other hand, we notice that the LPTC neurons also receive signals from medulla

neurons [31], [34]. If the medulla neurons which provide signals to the LPTC neurons, are laterally inhibited, the LPTC neurons would show strong size selectivity (this will be demonstrated in the experiment Section IV-B). This may conflict with the finding that the LPTCs do not exhibit size selectivity [17]–[19]. To satisfy both size selectivity of the STMDs and size insensitivity of the LPTCs, we infer that the second-order lateral inhibition mechanism may happen on the STMD pathway rather than medulla neurons in the implementation of our proposed neural network.

D. Lobula Layer

In the insect's visual system, the STMD neurons integrate signals from the medulla layer. They respond strongly to small target motion, but show weak or no response to wide-field motion (size selectivity) [2], [4], [5]. Besides, some STMDs exhibit strong responses to small target motion oriented along a preferred direction, but show weak or no response to opposite-direction motion (direction selectivity) [6], [7].

In the proposed neural network (DSTMD), a new correlation mechanism and a second-order lateral inhibition mechanism are introduced for direction and size selectivities, respectively. For comparison with the proposed neural network, the existing non-directionally selective ESTMD [9] is also presented in the following.

1) *ESTMD*: In the ESTMD, the output of an STMD neuron located at (x, y) , denoted by $\tilde{D}(x, y, t)$, is defined as,

$$\tilde{D}(x, y, t) = \tilde{S}^{\text{Tm3}}(x, y, t) \times \tilde{S}^{\text{Tm1}}_{D(n_3, \tau_3)}(x, y, t). \quad (24)$$

As we can see from (24), the output of an STMD neuron located at (x, y) is defined by the multiplication of the Tm1 and Tm3 outputs at the same position. Since medulla neural signals from at least two different positions are needed for detecting motion direction [35], the ESTMD is able to detect the presence of target motion, but not the target's motion direction.

2) *DSTMD*: In the DSTMD, the correlation output of an STMD neuron located at (x, y) with a preferred motion direction θ , denoted by $D(x, y, t; \theta)$, is defined as,

$$\begin{aligned} D(x, y, t; \theta) = & S^{\text{Tm3}}(x, y, t) \times \left\{ S^{\text{Tm1}}_{D(n_5, \tau_5)}(x, y, t) \right. \\ & \left. + S^{\text{Mi1}}_{D(n_4, \tau_4)}(x', y', t) \right\} \\ & \times S^{\text{Tm1}}_{D(n_6, \tau_6)}(x', y', t) \end{aligned} \quad (25)$$

where

$$\begin{aligned} x' &= x + \alpha_1 \cos \theta \\ y' &= y + \alpha_1 \sin \theta \end{aligned} \quad (26)$$

and α_1 is a constant, $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$.

As we can see from (25), four medulla neural signals from two different positions, i.e., (x, y) and (x', y') , are used to define the output of an STMD neuron located at (x, y) (see Fig. 2, two multipliers and one adder). These four medulla neural signals include the outputs of the Tm1 and Tm3 located at position (x, y) , i.e., $S^{\text{Tm1}}_{D(n_5, \tau_5)}(x, y, t)$ and $S^{\text{Tm3}}(x, y, t)$, the outputs of the Tm1 and Mi1 located at position (x', y') , i.e.,

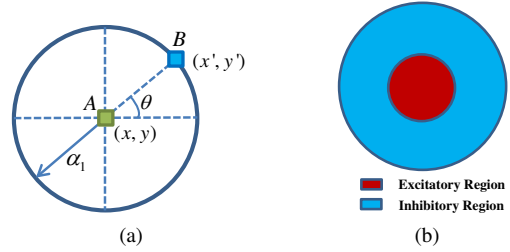


Fig. 5. (a) Schematic illustration of relative position between A (x, y) and B (x', y') . α_1 is the distance between A and B while θ is the angle between line segment AB and the horizontal line. (b) Schematic illustration of excitatory and inhibitory regions of the second-order lateral inhibition mechanism.

$S^{\text{Tm1}}_{D(n_6, \tau_6)}(x', y', t)$ and $S^{\text{Mi1}}_{D(n_4, \tau_4)}(x', y', t)$ (the full derivation of (25) is shown in the supplementary materials). The schematic illustration of relative position between (x, y) and (x', y') is presented in Fig. 5(a). For a given position (x, y) , we can choose a series of (x', y') , corresponding to different directions θ . Thus, a series of correlation outputs $D(x, y, t; \theta)$ with different preferred motion directions θ can be defined. For a given direction θ_0 , $D(x, y, t; \theta_0)$ gives the strongest output to small target motion oriented along direction θ_0 , with weak or no outputs to motion oriented along other directions. That is, $D(x, y, t; \theta)$ shows direction selectivity.

After the signal correlation, the DSTMD implements the second-order lateral inhibition mechanism on $D(x, y, t; \theta)$ for size selectivity. That is,

$$D_I(x, y, t; \theta) = \left[\iint D(u, v, t; \theta) W_2(x - u, y - v) dudv \right]^+ \quad (27)$$

where $D_I(x, y, t; \theta)$ is the signal after lateral inhibition and $[x]^+$ denotes $\max(x, 0)$, $W_2(x, y)$ is defined in (20).

The schematic illustration of inhibition kernel $W_2(x, y)$ is shown in Fig. 5(b). As can be seen, the inhibition kernel $W_2(x, y)$ contains two components, i.e., excitatory and inhibitory regions. For the kernel $W_2(x, y)$, its surround inhibition is set as three times as strong as the center excitation. In this case, once the target's size exceeds the excitatory region, it will receive strong inhibition. When the target is smaller than the excitatory region, the amount of excitation will increase as the rise of target size. That is, the DSTMD prefers the target whose size is equal to the excitatory region and exhibits size selectivity.

Following the second-order lateral inhibition mechanism, the DSTMD inhibits model output $D_I(x, y, t; \theta)$ at directions more than 45° apart by convolving $D_I(x, y, t; \theta)$ with an inhibition kernel $W_3(\theta)$. That is,

$$E(x, y, t; \theta) = \left[\int D_I(x, y, t; \varphi) W_3(\theta - \varphi) d\varphi \right]^+ \quad (28)$$

where $[x]^+$ denotes $\max(x, 0)$ and $W_3(\theta)$ is defined as

$$W_3(\theta) = G_{\sigma_6}(\theta) - G_{\sigma_7}(\theta). \quad (29)$$

where $G_\sigma(x, y)$ is a Gaussian function.

In the DSTMD, $E(x, y, t; \theta)$ is regarded as the output of the STMD neurons.

TABLE I
PARAMETERS OF THE DSTMD AND ESTMD

Eq.	Parameters
(1)	$\sigma_1 = 1$
(4)	$n_1 = 2, \tau_1 = 3, n_2 = 6, \tau_2 = 9$
(8), (9)	$\sigma_2 = 1.5, \sigma_3 = 3$
(10), (11)	$\lambda_1 = 3, \lambda_2 = 9$
(20)	$A = 1, B = 3$
(21)	$\sigma_4 = 1.5, \sigma_5 = 3, e = 1, \rho = 0$
(24)	$n_3 = 5, \tau_3 = 25$
(25)	$n_4 = 3, \tau_4 = 15, n_5 = 5, \tau_5 = 25, n_6 = 8, \tau_6 = 40$
(26)	$\alpha_1 = 3$
(29)	$\sigma_6 = 1.5, \sigma_7 = 3$

E. Motion Direction Estimation

In the insect's visual system, the STMD neurons are believed to be upstream of target selective descending neurons (TSDNs) [4], [6], [8]. Further biological research [8] found that eight pairs of the TSDNs are able to encode motion direction of targets by a population vector algorithm.

In the proposed neural network, we estimate motion directions of targets by populating the model output $E(x, y, t; \theta)$ along different directions θ . That is,

$$MD(t) = \sum_{(x,y) \in \text{Target}} \sum_{\theta} (E(x, y, t; \theta) \cos \theta, E(x, y, t; \theta) \sin \theta) \quad (30)$$

where $MD(t)$ denotes the motion direction of the small target at time t , $(x, y) \in \text{Target}$ stands for the position of the STMD neurons which respond to the small target motion.

F. Parameter Setting

Parameters of the proposed neural network (DSTMD) and ESTMD are given in Table I. These parameters are tuned manually based on empirical experiences and will not be changed in the following experiments unless stated.

The proposed neural network is written in Matlab (The MathWorks, Inc., Natick, MA). The computer used in the experiments is a PC with one 2.50 Ghz CPU (Core i7 4710MQ) and windows 7 operating system. The source code can be found at <https://github.com/wanghongxin/DSTMD>.

IV. RESULTS AND DISCUSSIONS

The proposed neural network is tested on image sequences produced by Vision Egg [36]. The Vision Egg is an open-source programming library that allows scientists to produce arbitrary visual stimuli (<http://visionegg.org/>). Such stimuli involve traditional stimuli such as sinusoidal gratings, or may be more complex, 3-D, and naturalistic scenes. The image sequences used in this paper can be divided into two categories depending on background types. The first category contains image sequences showing small target motion against white backgrounds. This category is used to test the basic properties of the proposed neural network, such as tuning properties (see Sections IV-B and IV-C), direction selectivity (see Section IV-D). The other category contains image sequences showing

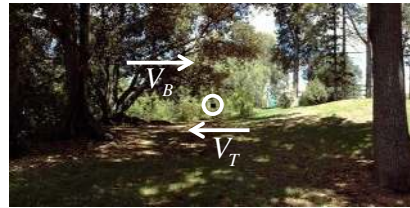


Fig. 6. Representative frame of the input image sequence. A small rectangle highlighted by the white circle, is moving against the cluttered background. This rectangle whose size and luminance are set as 5×5 pixels and 0, is the small target needed to be detected. The arrows V_T and V_B denote the motion directions of the small target and the background, respectively. The velocities of the small target and the background are all set as 250 pixel/s.

small target motion against naturally cluttered backgrounds. This category is used to test the detection performance of the proposed neural network in complex backgrounds (see Section IV-A and IV-E). All image sequences can be reproduced by the Vision Egg with the same parameters (given before each experiment). The video images are 500 (in horizontal) by 250 (in vertical) pixels and temporal sampling frequency is set as 1000 Hz.

A. Contribution of Various Neurons

To evaluate the characteristics of the neurons in the proposed neural network, we observe and analyze their outputs. For an input image sequence $I(x, y, t)$, where $x \in [0, 500]$ pixel, $y \in [0, 250]$ pixel, $t \in [0, 1000]$ ms (see Fig. 6), we first fix y and t as $y_0 = 125$ pixel and $t_0 = 1000$ ms, then illustrate $I(x, y_0, t_0)$ with respect to x in Fig. 7(a). Similarly, the outputs of other neurons are presented in the subplots below.

Fig. 7(a)-(c) shows the input luminance signal $I(x, y_0, t_0)$, ommatidium output $P(x, y_0, t_0)$ and LMC output $L(x, y_0, t_0)$, respectively. Compared to the input signal, the ommatidium output demonstrates little difference and is just slightly smoothed. This is because the ommatidium is modeled as a spatial Gaussian filter to smooth the input luminance signals. The LMC output displays significant difference from the ommatidium output. More precisely, the LMC output becomes the band-pass-filtered version of the ommatidium output. From the other perspective, the LMC output reveals the luminance changes of pixels, where the positive values correspond to luminance increase while the negative values suggest luminance decrease.

Fig. 7(d) and (e) illustrates the outputs of the Tm3 and Tm2 modeled by the DSTMD and ESTMD, respectively. Compared to Fig. 7(d), the outputs of the Tm3 and Tm2 are largely suppressed in Fig. 7(e). This is because the ESTMD uses the laterally inhibited ON and OFF signals to define the outputs of the Tm3 and Tm2 (see (18) and (19)), while the DSTMD utilizes the ON and OFF signals directly (see (12) and (13)). Fig. 7(f) and (g) demonstrate the medulla signals used for the signal correlation in the DSTMD and ESTMD, respectively. As can be seen, four medulla signals are used for the signal correlation in the DSTMD whereas only two medulla signals are utilized in the ESTMD.

Fig. 7(h) and (i) displays the outputs of the proposed DSTMD model and the existing non-directionally selective

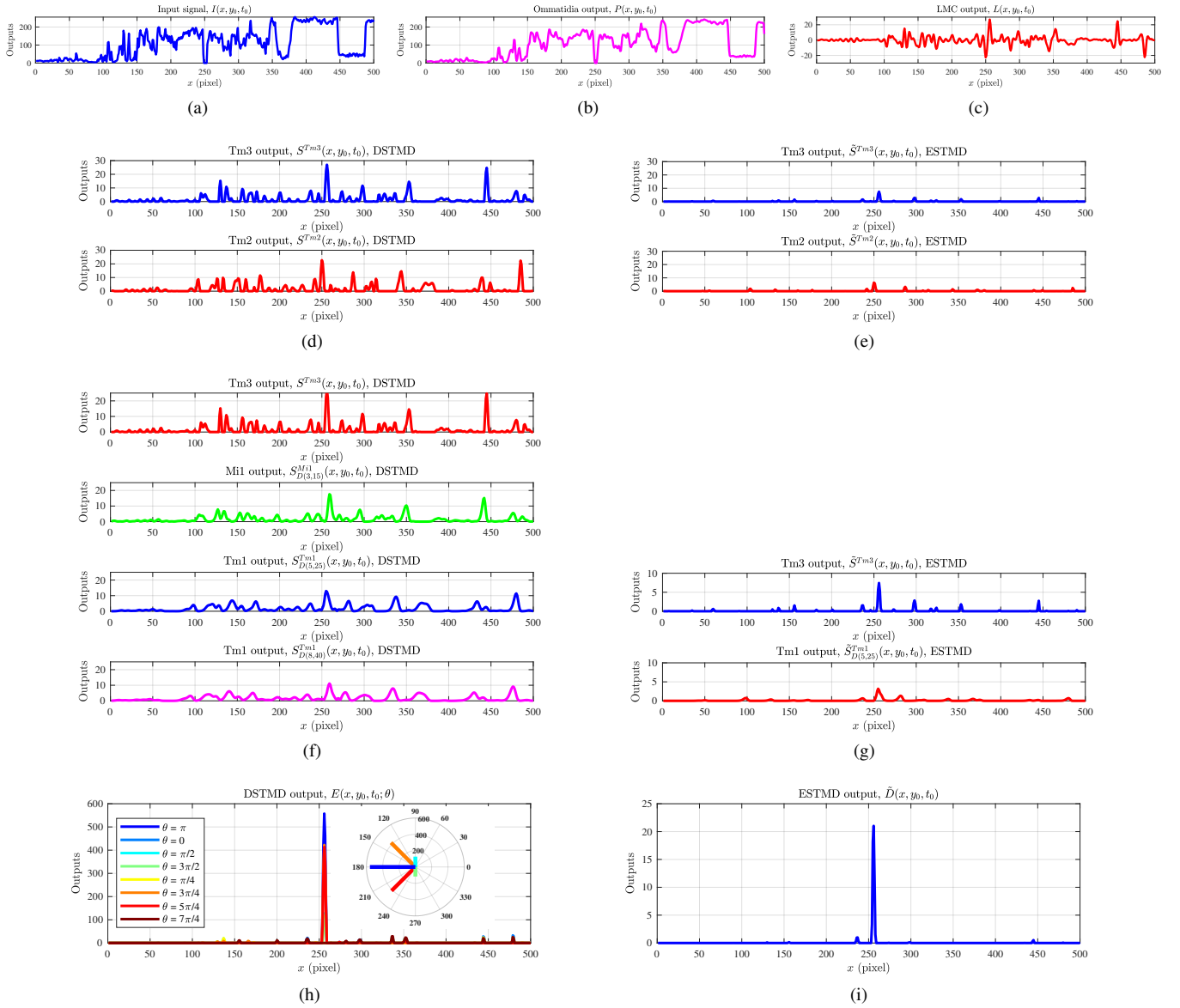


Fig. 7. Outputs of various neurons in the DSTMD and ESTMD models where $x \in [0, 500]$ pixel, $y_0 = 125$ pixel, $t_0 = 1000$ ms. In each subplot, the horizontal axis denotes the location of neurons (x coordinate) while the vertical axis represents neural outputs. (a) Input signal $I(x, y_0, t_0)$. (b) Ommatidium output $P(x, y_0, t_0)$. (c) LMC output $L(x, y_0, t_0)$. (d) Outputs of the Tm3 and Tm2 modeled by the DSTMD, i.e., $S^{Tm3}(x, y_0, t_0)$ and $S^{Tm2}(x, y_0, t_0)$. (e) Outputs of the Tm3 and Tm2 modeled by the ESTMD, i.e., $\tilde{S}^{Tm3}(x, y_0, t_0)$ and $\tilde{S}^{Tm2}(x, y_0, t_0)$. (f) Medulla neural outputs used for the signal correlation in the DSTMD, i.e., $S^{Tm3}(x, y_0, t_0)$, $S_{D(3,15)}^{M1}(x, y_0, t_0)$, $S_{D(5,25)}^{Tm1}(x, y_0, t_0)$ and $S_{D(8,40)}^{Tm1}(x, y_0, t_0)$. (g) Medulla neural outputs used for the signal correlation in the ESTMD, i.e., $\tilde{S}^{Tm3}(x, y_0, t_0)$ and $\tilde{S}_{D(5,25)}^{Tm1}(x, y_0, t_0)$. (h) DSTMD output $E(x, y_0, t_0; \theta)$. In this subplot, the DSTMD outputs to the small target are further shown in the polar coordinate system, where the angular coordinate denotes the preferred motion direction θ while the radial coordinate stands for the model output along this preferred direction. (i) ESTMD output $\tilde{D}(x, y_0, t_0)$.

ESTMD model, respectively. From these two subplots, we can see that both the DSTMD and ESTMD show the strongest response at $x = 250$ which is the location of the small moving target. At the other positions, both models exhibit much weaker or even no response. For example, both models demonstrate little response to the tree trunk located between $x = 450$ and $x = 480$, which is regarded as a large object. The above results indicate that both the DSTMD and ESTMD are only interested in small target motion.

Comparing Fig. 7(h) with Fig. 7(i), we can find that the major difference between the DSTMD and ESTMD is direction selectivity. More precisely, in Fig. 7(h), the DSTMD has

eight outputs $E(x, y_0, t_0; \theta)$ corresponding to eight preferred directions θ , $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$. However, in Fig. 7(i), the ESTMD only has one output $\tilde{D}(x, y_0, t_0)$ lacking of direction information. To clearly show direction selectivity, the DSTMD outputs to the small target are illustrated in polar coordinate (see the right part of Fig. 7(h)). As can be seen, the DSTMD exhibits the strongest output along direction $\theta = \pi$ which is consistent with the motion direction of the small target. The other seven outputs of the DSTMD decrease as the corresponding direction θ deviates from the small target motion direction.

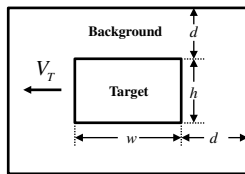


Fig. 8. External rectangle and neighboring background rectangle of a small target. The arrow V_T denotes the motion direction of the target. w represents target width while h stands for target height.

B. Tuning Properties

We test four basic properties of the proposed neural network, including Weber Contrast sensitivity, velocity selectivity, width selectivity and height selectivity. These four properties are basic properties of the STMD neurons and are used to distinguish the STMD neurons in biology [2], [5], [7]. Here, Weber Contrast sensitivity refers to that the STMD neural response increases with the increase of Weber Contrast. Velocity selectivity refers to that the STMD neurons show the strongest response to a specific velocity (optimal velocity). Above or below this optimal velocity will result in the significant decrease of neural responses. Width selectivity and height selectivity are similar to velocity selectivity.

We first give definitions of Weber Contrast, width and height. As it is shown in Fig. 8, width represents the target length extended parallel to the motion direction while height denotes the target length extended orthogonal to the motion direction. If the size of a target is $w \times h$, the size of its background rectangle is $(w + 2d) \times (h + 2d)$, where d is a constant which equals to 10 pixels in this paper. Weber Contrast is defined by the following equation,

$$\text{Weber Contrast} = \frac{|\mu_t - \mu_b|}{255} \quad (31)$$

where μ_t is the average pixel value of the target, μ_b is the average pixel value in neighboring area around the target.

We perform four experiments to illustrate four basic properties of the proposed neural network. In these four experiments, image sequences which display a small target moving against the white background, are used as the model input. The initial parameters of the small target including luminance, velocity, width and height, are set as 0, 250 pixel/second, 5 pixels and 5 pixels, respectively. In each experiment, we change one of four target parameters while fix other three parameters, then record corresponding model outputs. The recorded tuning curves are displayed in Fig. 9.

As it is shown in Fig. 9(a), the outputs of the DSTMD and ESTMD increase with the increase of Weber Contrast, until reach maximum at Weber Contrast = 1. This reveals that the DSTMD and ESTMD exhibit Weber Contrast sensitivity. In Fig. 9(b), the outputs of the DSTMD and ESTMD all peak at velocity = 300 pixel/s and decrease significantly when the target velocity is above or below 300 pixel/s. This suggests that the DSTMD and ESTMD have a preferred velocity and exhibit velocity selectivity. Similar variation trends can be seen in Fig. 9(c) and (d) which reveal the width selectivity and height selectivity of the DSTMD, respectively.

In the following, we carry out an experiment to demonstrate the hypothesis raised in Section III-C. The hypothesis is that if the medulla neurons which provide signals to the LPTC neurons [17]–[19], are laterally inhibited, the LPTC neurons would show strong size selectivity. In order to demonstrate this point, we first adopt TQD model [26], [27] to simulate the LPTC neurons. Then we use the medulla neuron modelling methods of DSTMD and ESTMD to simulate medulla neurons, respectively. For TQD which receives signals from medulla neurons modeled by DSTMD, we denote it as TQD. For TQD which receives signals from medulla neurons modeled by ESTMD, we denote it as TQD(LI). The only difference between the TQD and TQD(LI) is that medulla neurons providing signals to the TQD(LI), are laterally inhibited. Finally, we test the four basic properties of the TQD and TQD(LI). The recorded tuning curves are presented in Fig. 10.

As it can be seen from Fig. 10(a), (b) and (c), the TQD and TQD (LI) display little difference. They all exhibit Weber Contrast sensitivity and velocity selectivity, but do not show the width selectivity. In Fig. 10(d), although both TQD and TQD (LI) have a local maximum at height = 5, they show differences with increasing height. As the continuous increase of the height, the output of the TQD firstly has a slight drop and finally tends to be stable around 0.9. In contrast, the output of the TQD (LI) decreases significantly and finally tends to be stable around 0.05. Above results indicate that the TQD(LI) exhibits height selectivity. This contradicts with the biological finding that the LPTC neurons are not size selective [17]–[19]. To avoid conflict with the biological finding on the LPTC neurons, we adopt the new medulla neuron modeling method and implement the second-order lateral inhibition mechanism on the STMD neuron pathways.

C. Parameter Sensitivity

In the last section, we have demonstrated that the proposed neural network shows four basic properties, i.e., Weber Contrast sensitivity, velocity selectivity, width selectivity and height selectivity. In this section, we further evaluate the impacts of three sets of parameters, including (n_4, τ_4) , (n_5, τ_5) and (σ_4, σ_5) , on the four basic properties. These three sets of parameters are defined in (21) and (25).

We conduct three experiments to assess the effects of these three sets of parameters, respectively. In each experiment, we change one set of parameters while keep other two sets of parameters at their initially assigned value [see Table I], then record corresponding model outputs. In the first experiment, (n_4, τ_4) is set as (1, 5), (2, 10), (3, 15), (4, 20), (5, 25), (6, 30). In the second experiment, (n_5, τ_5) is set as (3, 15), (4, 20), (5, 25), (6, 30), (7, 35), (8, 40). In the third experiment, (σ_4, σ_5) is set as (1.0, 2.0), (1.5, 3.0), (2.3, 4.6), (2.8, 5.6), (3.7, 7.4). The recorded tuning curves of the proposed neural network under different parameter settings, are presented in Fig. 11–13.

In the first and second experiment, we illustrate that the parameter (n_4, τ_4) and (n_5, τ_5) have large impact on the velocity selectivity and width selectivity, but show little effect on the Weber Contrast sensitivity and height selectivity. More

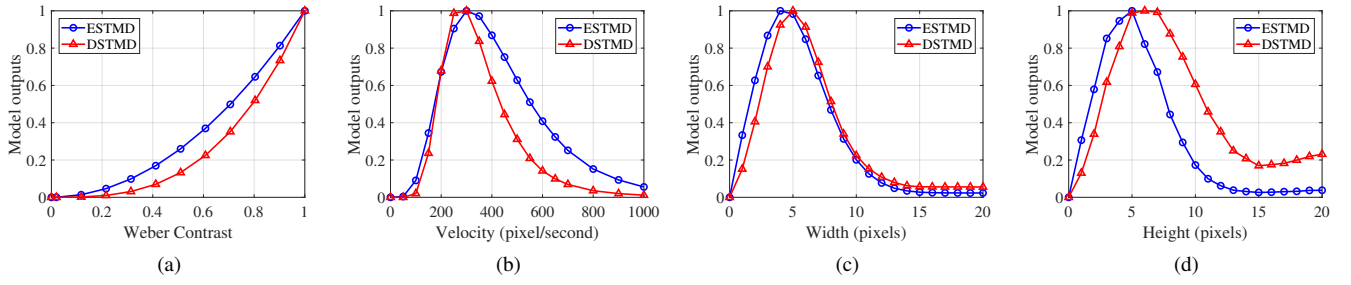


Fig. 9. Tuning properties of the proposed neural network (DSTMD) and ESTMD. In each subplot, the horizontal axis represents one of target parameters (Weber Contrast, velocity, width, and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

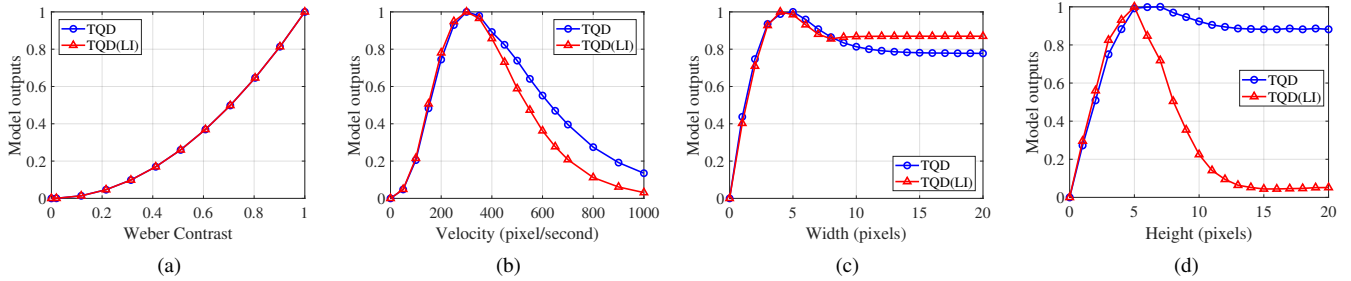


Fig. 10. Tuning properties of the TQD and TQD(LI). In each subplot, the horizontal axis represents one of target parameters (Weber Contrast, velocity, width, and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

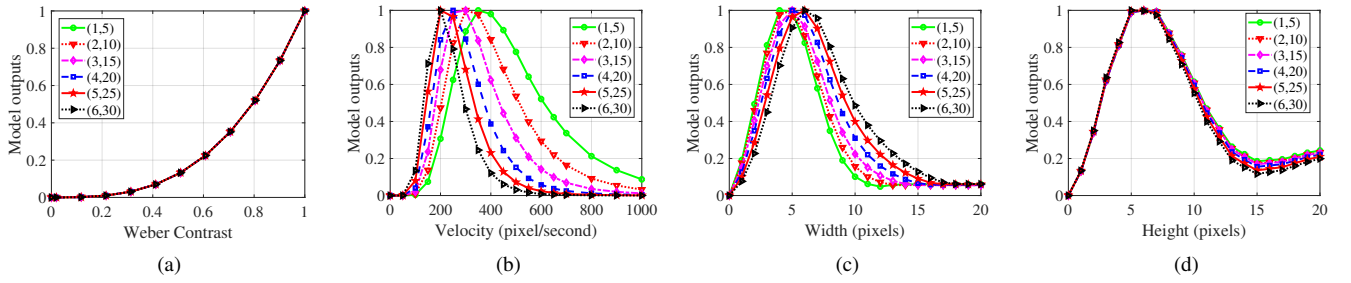


Fig. 11. Tuning properties of the proposed neural network under different parameter (n_4, τ_4) . In this experiment, (n_4, τ_4) is set as (1, 5), (2, 10), (3, 15), (4, 20), (5, 25), (6, 30) while the other parameters are fixed. In each subplot, the horizontal axis represents one of the target parameters (Weber Contrast, velocity, width, and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

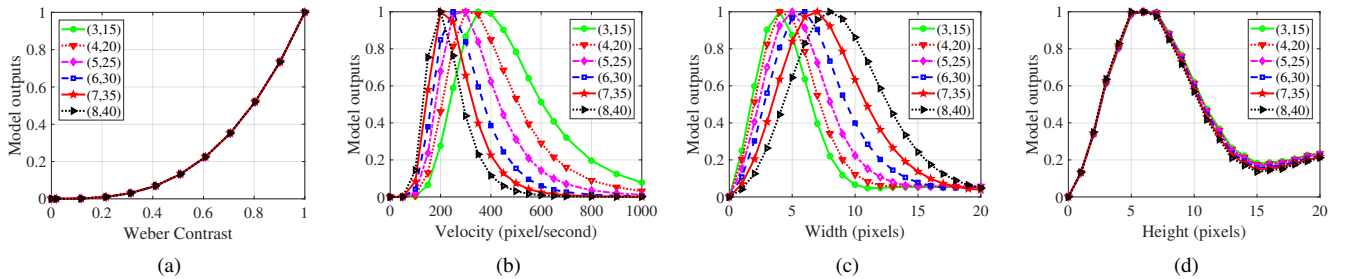


Fig. 12. Tuning properties of the proposed neural network under different parameter (n_5, τ_5) . In this experiment, (n_5, τ_5) is set as (3, 15), (4, 20), (5, 25), (6, 30), (7, 35), (8, 40) while the other parameters are fixed. In each subplot, the horizontal axis represents one of the target parameters (Weber Contrast, velocity, width, and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

precisely, from Fig. 11(a) and (d), we can see that the increase of (n_4, τ_4) have not induced any significant changes of the Weber Contrast tuning curve and the height tuning curve. However, with the increase of (n_4, τ_4) , as shown in Fig. 11(b)

and (c), the peak velocity decreases while the peak width increases. In Fig. 12, the parameter (n_5, τ_5) has similar effect with (n_4, τ_4) on the four basic properties.

The reasons for the above results are—in the proposed

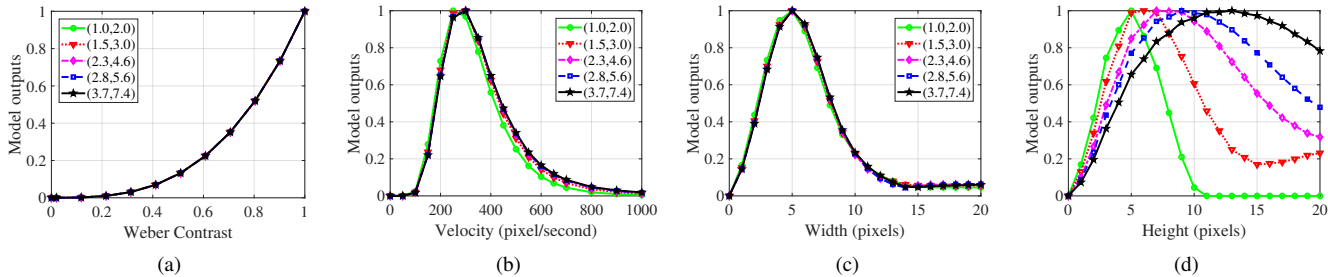


Fig. 13. Tuning properties of the proposed neural network under different parameter (σ_4, σ_5) . In this experiment, (σ_4, σ_5) is set as $(1.0, 2.0)$, $(1.5, 3.0)$, $(2.3, 4.6)$, $(2.8, 5.6)$, $(3.7, 7.4)$ while the other parameters are fixed. In each subplot, the horizontal axis represents one of the target parameters (Weber Contrast, velocity, width, and height) while the vertical axis denotes normalized model outputs. (a) Weber Contrast tuning curves. (b) Velocity tuning curves. (c) Width tuning curves. (d) Height tuning curves.

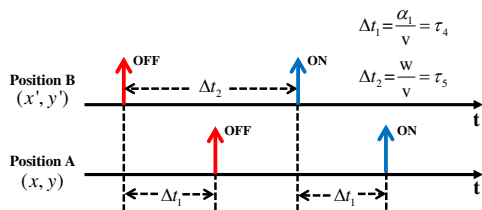


Fig. 14. Schematic illustration of the luminance changes of the position A and B when a dark target successively passes position B (x', y') and A (x, y) . The red arrow denotes luminance decrease signal (OFF signal) while the blue arrow represents luminance increase signal (ON signal). Let α_1 , w and v stand for the distance between position A and B, target width and velocity, respectively. Then we have $\Delta t_1 = \frac{\alpha_1}{v}$, $\Delta t_2 = \frac{w}{v}$.

neural network, τ_4 and τ_5 are positively correlated to $\frac{\alpha_1}{v}$ and $\frac{w}{v}$, respectively, where α_1 , v and w stand for the distance between position A and B, the peak velocity and the peak width, respectively. Once α_1 is given, the increase of τ_4 (or τ_5) will result in the decrease of the peak velocity v and the increase of the peak width w .

We further explain why τ_4 and τ_5 are positively correlated to $\frac{\alpha_1}{v}$ and $\frac{w}{v}$. In Fig. 14, we present the luminance changes of position A and B when a dark small target moves from B to A. In the equation (25), the DSTMD uses four medulla signals from position A (x, y) and B (x', y') to define the output of STMD neurons. Combining Fig. 14 with the equation (25), we point out that these four medulla signal are: 1) ON signal of position A (x, y) , corresponding to $S^{Tm3}(x, y, t)$; 2) ON signal of position B (x', y') with time delay order n_4 and time delay length τ_4 , corresponding to $S_{D(n_4, \tau_4)}^{M_{i1}}(x', y', t)$; 3) OFF signal of position A (x, y) with time delay order n_5 and time delay length τ_5 , corresponding to $S_{D(n_5, \tau_5)}^{Tm1}(x, y, t)$; 4) OFF signal of position B (x', y') with time delay order n_6 and time delay length τ_6 , corresponding to $S_{D(n_6, \tau_6)}^{Tm1}(x', y', t)$. In the DSTMD, we set τ_4 , τ_5 and τ_6 as Δt_1 , Δt_2 and $\Delta t_1 + \Delta t_2$, respectively. Since $\Delta t_1 = \frac{\alpha_1}{v}$ and $\Delta t_2 = \frac{w}{v}$, then we have $\tau_4 = \frac{\alpha_1}{v}$ and $\tau_5 = \frac{w}{v}$. That is, τ_4 and τ_5 are positively correlated to $\frac{\alpha_1}{v}$ and $\frac{w}{v}$, respectively.

In the third experiment, we demonstrate that the parameter (σ_4, σ_5) has large impact on the height selectivity, but shows little effect on the other three properties. As it can be seen from Fig. 13(a)-(c), the tuning curves have little changes with the

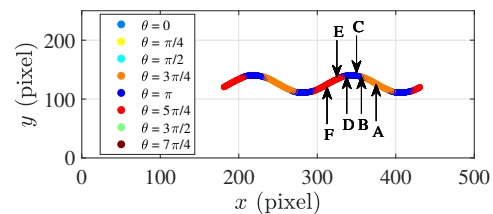


Fig. 15. Motion trace of the small target where color denotes the direction of the strongest output of the proposed neural network.

increase of (σ_4, σ_5) ; in contrast, the peak height of the height tuning curve increases, as presented in Fig. 13(d). Here, we point out that the peak height is positively correlated to the size of the excitatory region of the lateral inhibition mechanism (see Fig. 5(b)). In the proposed neural network, the size of the excitatory region is determined by σ_4 and σ_5 , where the higher (σ_4, σ_5) means the larger excitatory region, i.e., the larger peak height.

D. Direction Selectivity and Motion Direction Estimation

In this section, we illustrate how the proposed neural network encode motion directions of small targets. In the experiment, an image sequence which displays a small target moving against the white background, is used as the network input. The luminance and size of the small target are set as 0 and 5×5 pixels, respectively. The coordinate of the small target at time t is $(500 - 250 \cdot \frac{t+300}{1000}, 125 + 15 \cdot \sin(4\pi \frac{t+300}{1000}))$, $t \in [0, 1000]$ ms. Fig. 15 presents the motion trace of the small target. The motion direction of the small target varies between 142.98° and 217.01° when it moves along this motion trace.

We select six positions on the motion trace (A-F, in Fig. 15). The outputs of the DSTMD at these six positions are normalized, then shown in polar coordinate (see Fig. 16). In each subplot of Fig. 16, we can see that the smaller difference between the preferred direction θ and the actual motion direction (shown in Fig. 17), the stronger DSTMD output tuned to this direction θ . These directionally selective outputs are used to encode the motion direction of the small target by the population vector algorithm. Fig. 17 and Table II show the estimated motion direction and the actual motion direction at the six positions. As can be seen, the difference between the estimated direction and actual direction is smaller

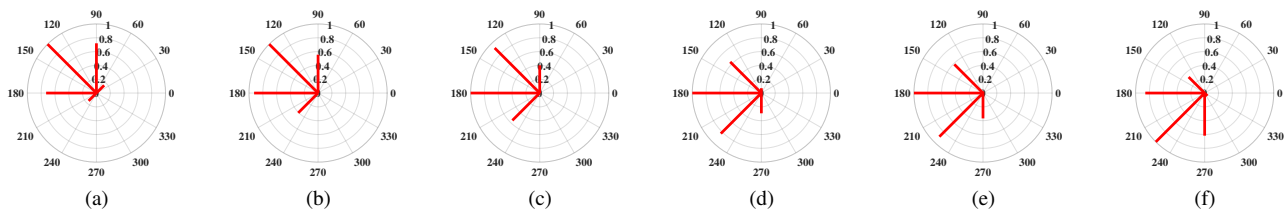


Fig. 16. (a)-(f) Normalized DSTMD outputs at the position A-F. In each subplot, the angular coordinate represents the preferred motion direction of the DSTMD while the radial coordinate denotes the strength of the DSTMD output tuned to this preferred direction.

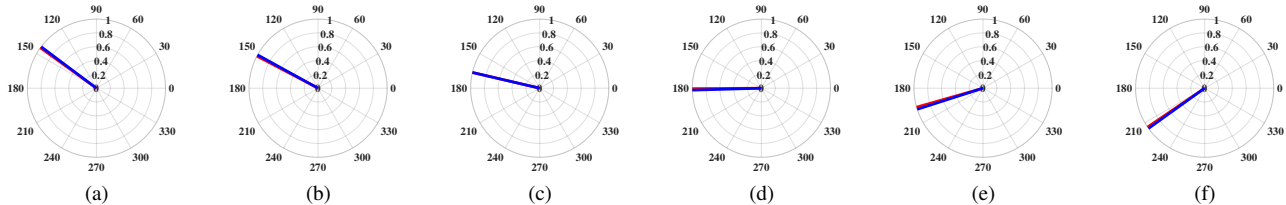


Fig. 17. (a)-(f) Estimated motion direction (red) and actual motion direction (blue) at the position A-F. In each subplot, the red line is highly overlapped with the blue line. That is, the estimated motion direction is quite close to the actual motion direction.

TABLE II
ESTIMATED MOTION DIRECTION AND ACTUAL MOTION DIRECTION AT THE SIX POSITIONS

Position	Estimated	Actual	Difference
A	144.25°	143.12°	1.13°
B	152.36°	151.21°	1.15°
C	166.83°	166.88°	0.05°
D	180.37°	181.63°	1.26°
E	195.93°	197.80°	1.87°
F	214.24°	215.53°	1.29°

than 2° at these six positions. We further estimate the motion direction of the small target at each position of the motion trace. The maximal difference between the estimated motion direction and actual motion direction is 3.17°. Above results indicate that the proposed neural network provides a good estimation for the motion direction of the small target.

E. Target Detection in Cluttered Backgrounds

In this section, we test the ability of the proposed neural network for detecting small targets against cluttered backgrounds. For a given detection threshold γ , if there is a position (x_0, y_0) , time t_0 and direction θ_0 which satisfy the DSTMD output $E(x_0, y_0, t_0; \theta_0) > \gamma$, then we believe that a small target is detected at position (x_0, y_0) and time t_0 . Two metrics are defined to evaluate the detection performance. That is,

$$D_R = \frac{\text{number of true detections}}{\text{number of actual targets}} \quad (32)$$

$$F_A = \frac{\text{number of false detections}}{\text{number of images}} \quad (33)$$

where D_R and F_A denote the detection rate and false alarm rate, respectively. The detected result is considered correct if the pixel distance between the ground truth and the result is within a threshold (5 pixels).

In the first three experiments, we investigate the influences of three target parameters (size, luminance and velocity) on

TABLE III
SETTINGS OF THE PARAMETERS INCLUDING TARGET LUMINANCE, SIZE AND HORIZONTAL VELOCITY FOR THE FIRST THREE EXPERIMENTS

	Luminance	Size	Velocity (V_T^x)
Experiment 1	0, 25, 50	5 × 5	250
Experiment 2	0	3 × 3, 5 × 5, 8 × 8	250
Experiment 3	0	5 × 5	200, 250, 350



Fig. 18. Representative frame of the input image sequence. The small target is highlighted by the white circle. The white arrow V_B denotes the motion direction of the background.

the detection performance. In each experiment, we change one of the target parameters while fix the other two parameters, then record the detection performance of the models under this parameter setting. The parameter settings of the first three experiments are shown in Table III. All input image sequences are produced using the same background image where a representative frame is given in Fig. 18. In all input image sequences, the background is moving from left to right and its velocity is set as 250 pixel/s. A small target is moving against the cluttered background, and its coordinate at time t is $(500 - V_T^x \cdot \frac{t+300}{1000}, 125 + 15 \cdot \sin(4\pi \frac{t+300}{1000}))$, $t \in [0, 1000]$ ms where V_T^x denotes the horizontal velocity. The receiver operating characteristic (ROC) curves of the three experiments with respect to target luminance, size and horizontal velocity V_T^x , are displayed in Fig. 19.

In Fig. 19(a), we can see that the lower target luminance is, the better ESTMD and DSTMD perform. This is because the decrease of target luminance can induce the increase of Weber Contrast (see Fig. S3 in the supplementary material). Note that the ESTMD and DSTMD all show Weber Contrast

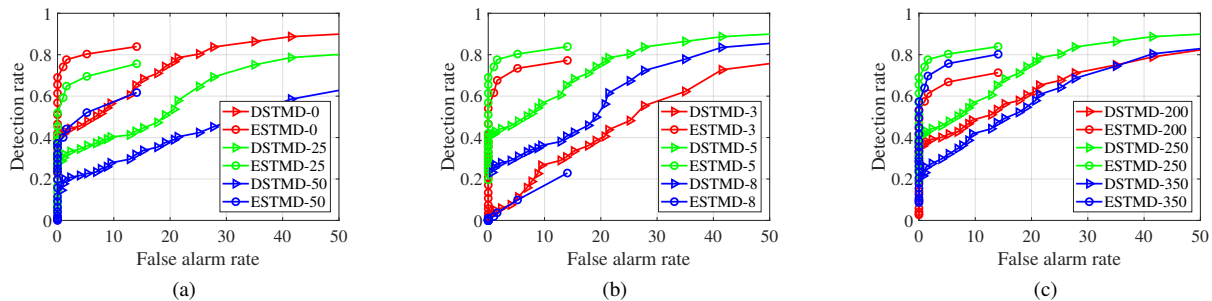


Fig. 19. ROC curves of the first three experiments with respect to target luminance, sizes and velocities. (a) Experiment 1, different target luminance. Legend 'ESTMD-0' and 'DSTMD-0' represent the ROC curves of the ESTMD and DSTMD when target luminance equals to 0, respectively. Similar explanations for other legends. (b) Experiment 2, different target sizes. Legend 'ESTMD-3' and 'DSTMD-3' represent the ROC curves of the ESTMD and DSTMD when target size equals to 3×3 pixels, respectively. Similar explanations for other legends. (c) Experiment 3, different horizontal velocities (V_T^x). Legend 'ESTMD-200' and 'DSTMD-200' represent the ROC curves of the ESTMD and DSTMD when the horizontal velocity V_T^x equals to 200 pixel/s, respectively. Similar explanations for other legends.

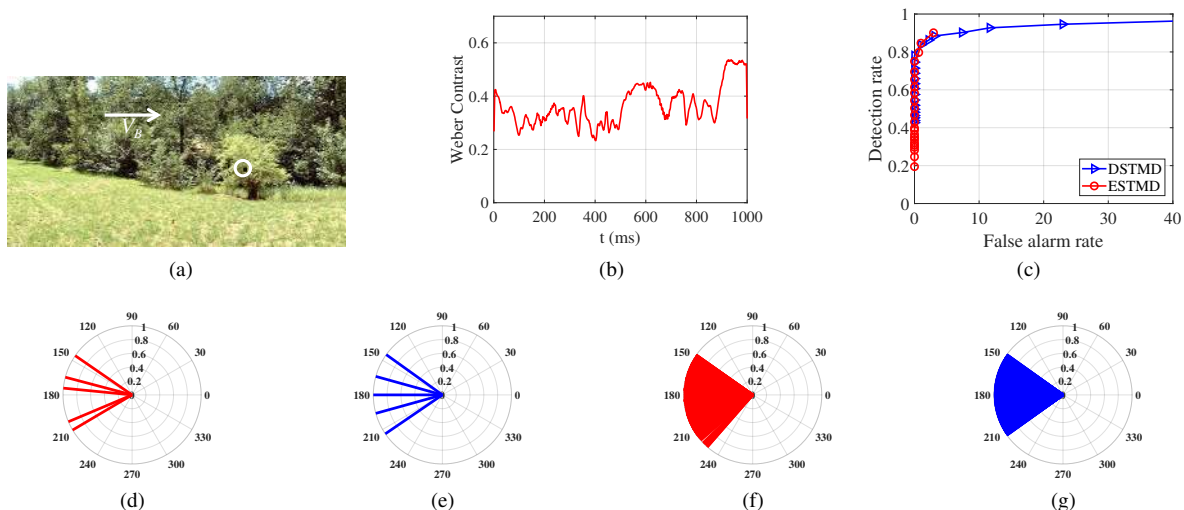


Fig. 20. Experiment 4. (a) Representative frame of the input image sequence. (b) Weber Contrast of the small target during time period $t \in [0, 1000]$ ms. (c) ROC curves of the ESTMD and DSTMD. (d) Motion directions detected by the DSTMD in the sample 510, 570, 600, 630, 700 frames. No motion direction detected by the ESTMD. (e) Actual motion directions in the sample 510, 570, 600, 630, 700 frames. (f) Motion directions detected by the DSTMD from the 500th to the 700th frame. (g) Actual motion directions from the 500th to the 700th frame.

sensitivity, so the higher Weber Contrast can elicit the stronger model output. From Fig. 19(b) and 19(c), we can see that when the false alarm rate is given, the target size of 5×5 (or the velocity of 250) has higher detection rate compared to the target size of 3×3 and 8×8 (or the velocity of 200 and 350). This is because the ESTMD and DSTMD all exhibit size and velocity selectivities. They show the strongest output to the target whose size (or velocity) equals to 5×5 pixels (or 250 pixel/s), but weaker outputs to the object whose size (or velocity) is higher or lower than 5×5 pixels (or 250 pixel/s).

In the fourth and fifth experiment, we evaluate the performance of the proposed neural network in different backgrounds. Two input image sequences with different backgrounds are displayed in Fig. 20(a) and Fig. 21(a), respectively. In these two image sequences, the backgrounds are all moving from left to right and their velocities are set as 250 pixel/s. A small target whose luminance and size are set as 0 and 5×5 pixels, is moving against the cluttered backgrounds. The coordinate of the small target at time t is $(500 - 250 \cdot \frac{t+300}{1000}, 125 + 15 \cdot \sin(4\pi \frac{t+300}{1000}))$, $t \in [0, 1000]$ ms.

Fig. 20(c) and Fig. 21(c) demonstrate the ROC curves for

the two image sequences, respectively. As can be seen, the detection rates of the DSTMD (or ESTMD) in Fig. 21(c) are much lower than those in Fig. 20(c). There are two reasons for the above result: 1) the background in Fig. 21(a) is more cluttered compared to Fig. 20(a), which means that it contains more small-target-like background features and 2) the Weber Contrast in Fig. 21(b) is much lower than that in Fig. 20(b), suggesting that the models exhibit much weaker outputs to the small target in the fifth experiment.

Fig. 20(d) displays the motion directions detected by the DSTMD in the sample 510, 570, 600, 630, 700 frames while Fig. 20(f) illustrates the motion directions detected by the DSTMD from the 500th to the 700th frame. As it is shown, these detected motion directions are quite close to the actual motion directions in Fig. 20(e) and 20(g). No motion direction is detected by the ESTMD, because it is not directionally selective. Similar results can be seen in Fig. 21(d)-(g).

The proposed neural network is further tested on a set of real videos. The experimental results are presented in the supplementary materials, due to the page limit.

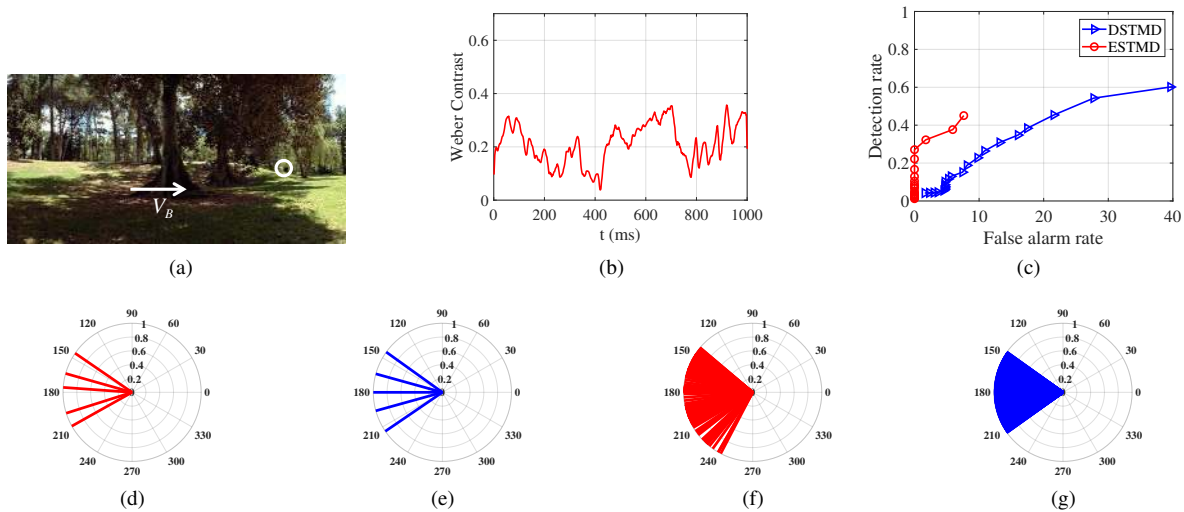


Fig. 21. Experiment 5. (a) Representative frame of the input image sequence. (b) Weber Contrast of the small target during time period $t \in [0, 1000]$ ms. (c) ROC curves of the ESTMD and DSTMD. (d) Motion directions detected by the DSTMD in the sample 510, 570, 600, 630, 700 frames. No motion direction detected by the ESTMD. (e) Actual motion directions in the sample 510, 570, 600, 630, 700 frames. (f) Motion directions detected by the DSTMD from the 500th to the 700th frame. (g) Actual motion directions from the 500th to the 700th frame.

V. FURTHER DISCUSSIONS

In the above sections, the presented neural network (DSTMD) demonstrated a reliable ability to detect small targets and motion directions against complex backgrounds. Nowadays, for vision-based mobile robots, their visual sensors are becoming more reliable while computation ability is more powerful. These make it possible for mobile robots, such as unmanned aerial vehicle (UAV), equipped with the presented neural network to detect small moving targets in the distance in the real world.

In the insects' visual system, numerous neurons work together to extract different cues from the real world. For example, the LMCs extract motion information while the amacrine cells capture contrast information from input visual signals [37], [38]. Integrating these two types of information may contribute to the improvement of detection performance of the STMD neurons in cluttered backgrounds. In the future, the cooperation of these specialized neurons needs to be taken into consideration.

A number of bio-inspired neural networks based on firing-rate methods, spiking neural networks or convolutional neural networks [39]–[41], have been used for target detection, tracking and navigation. Although these neural networks perform well, they cannot distinguish small target motion from large object motion. Detecting target motion is relatively easy, but distinguishing different target motion in terms of the targets' sizes is more challenging and difficult. For example, a naturally cluttered background always contains small targets such as insects, and large objects such as bushes, trees or rocks. Due to the camera motion, these large objects are moving with the background. In this case, the above-mentioned neural networks can detect both small and large object motion, but cannot distinguish them.

In engineering, small target motion detection can be performed by infrared detection methods [42]. However, these infrared methods always require significant temperature differences between objects of interest (such as rockets and jets) and

the background. This largely limits their application, because such significant temperature difference is rare in the natural world. Different from the infrared methods, the presented neural network uses normal images as input and provides a vision-based method for small moving target detection.

VI. CONCLUSION

In this paper, we proposed a visual neural network (DSTMD) to simulate the directionally selective STMD neurons. Direction selectivity is obtained by correlating signals from two positions while size selectivity is introduced by the second-order lateral inhibition mechanism. Motion directions of detected targets are estimated by the population vector algorithm. Systematic experiments showed that the presented STMD-based neural network can detect not only small moving targets, but also motion directions against complex backgrounds. In the future work, various visual neurons which extract different cues simultaneously, will be integrated together to improve detection performance.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments. They would also like to thank M. Daniels for proofreading this manuscript.

REFERENCES

- [1] S. Yue and F. C. Rind, "Collision detection in complex dynamic scenes using an lgmd-based visual neural network with feature enhancement," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 705–716, May 2006.
- [2] K. Nordström, P. D. Barnett, and D. C. O'Carroll, "Insect detection of small targets moving in visual clutter," *PLoS Biol.*, vol. 4, no. 3, p. e54, Feb. 2006.
- [3] R. Olberg, A. Worthington, and K. Venator, "Prey pursuit and interception in dragonflies," *J. Comp. Physiol. A*, vol. 186, no. 2, pp. 155–162, Feb. 2000.
- [4] K. Nordström, "Neural specializations for small target detection in insects," *Curr. Opin. Neurobiol.*, vol. 22, no. 2, pp. 272–278, Apr. 2012.
- [5] K. Nordström and D. C. O'Carroll, "Small object detection neurons in female hoverflies," *Proc. R. Soc. Lond., B, Biol. Sci.*, vol. 273, no. 1591, pp. 1211–1216, May 2006.

- [6] D. O'Carroll, "Feature-detecting neurons in dragonflies," *Nature*, vol. 362, no. 6420, pp. 541–543, Apr. 1993.
- [7] P. D. Barnett, K. Nordström, and D. C. O'Carroll, "Retinotopic organization of small-field-target-detecting neurons in the insect visual system," *Curr. Biol.*, vol. 17, no. 7, pp. 569–578, Apr. 2007.
- [8] P. T. Gonzalez-Bellido, H. Peng, J. Yang, A. P. Georgopoulos, and R. M. Olberg, "Eight pairs of descending visual neurons in the dragonfly give wing motor centers accurate population vector of prey direction," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 2, pp. 696–701, Jan. 2013.
- [9] 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.
- [10] 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, Dec. 2013.
- [11] Z. M. Bagheri, B. S. Cazzolato, S. Grainger, D. C. O'Carroll, and S. D. Wiederman, "An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments," *J. Neural Eng.*, vol. 14, no. 4, p. 046030, Jul. 2017.
- [12] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Performance of an insect-inspired target tracker in natural conditions," *Bioinspir. & Biomim.*, vol. 12, no. 2, p. 025006, Feb. 2017.
- [13] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Properties of neuronal facilitation that improve target tracking in natural pursuit simulations," *J. R. Soc. Interface*, vol. 12, no. 108, p. 20150083, Jun. 2015.
- [14] S. Judge and F. Rind, "The locust dcmd, a movement-detecting neurone tightly tuned to collision trajectories," *J. Exp. Biol.*, vol. 200, no. 16, pp. 2209–2216, Aug. 1997.
- [15] F. C. Rind and P. J. Simmons, "Orthopteran dcmd neuron: a reevaluation of responses to moving objects. i. selective responses to approaching objects," *J. Neurophysiol.*, vol. 68, no. 5, pp. 1654–1666, Nov. 1992.
- [16] P. J. Simmons and F. C. Rind, "Responses to object approach by a wide field visual neurone, the lgmd2 of the locust: characterization and image cues," *J. Comp. Physiol. A*, vol. 180, no. 3, pp. 203–214, Feb. 1997.
- [17] Y.-J. Lee, H. O. Jönsson, and K. Nordström, "Spatio-temporal dynamics of impulse responses to figure motion in optic flow neurons," *PLoS One*, vol. 10, no. 5, pp. 1–16, May 2015.
- [18] A. Borst and M. Egelhaaf, "Direction selectivity of blowfly motion-sensitive neurons is computed in a two-stage process," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 87, no. 23, pp. 9363–9367, Dec. 1990.
- [19] A. Borst, M. Egelhaaf, and J. Haag, "Mechanisms of dendritic integration underlying gain control in fly motion-sensitive interneurons," *J. Comput. Neurosci.*, vol. 2, no. 1, pp. 5–18, Mar. 1995.
- [20] F. C. Rind and D. Bramwell, "Neural network based on the input organization of an identified neuron signaling impending collision," *J. Neurophysiol.*, vol. 75, no. 3, pp. 967–985, Mar. 1996.
- [21] S. Yue and F. C. Rind, "Redundant neural vision systems-competing for collision recognition roles," *IEEE Trans. Auton. Mental Develop.*, vol. 5, no. 2, pp. 173–186, Apr. 2013.
- [22] S. Yue, F. C. Rind, M. S. Keil, J. Cuadri, and R. Stafford, "A bio-inspired visual collision detection mechanism for cars: Optimisation of a model of a locust neuron to a novel environment," *Neurocomputing*, vol. 69, no. 13, pp. 1591–1598, Feb. 2006.
- [23] Q. Fu, C. Hu, and S. Yue, "Bio-inspired collision detector with enhanced selectivity for ground robotic vision system," in *Proc. Brit. Mach. Vis. Conf.*, York, U.K., 2016.
- [24] C. Hu, F. Arvin, C. Xiong, and S. Yue, "A bio-inspired embedded vision system for autonomous micro-robots: the lgmd case," *IEEE Trans. Cogn. Develop. Syst.*, vol. 9, no. 3, pp. 241–254, Sep. 2016.
- [25] B. Hassenstein and W. Reichardt, "Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenbewertung bei der bewegungsperzeption des rüsselkäfers chlorophanus," *Zeitschrift für Naturforschung B*, vol. 11, no. 9-10, pp. 513–524, Oct. 1956.
- [26] N. Franceschini, A. Riehle, and A. Le Nestour, "Directionally selective motion detection by insect neurons," in *Facets of vision*. Berlin, Heidelberg: Springer, 1989, pp. 360–390.
- [27] 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.
- [28] D. A. Clark, L. Bursztyrn, 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.
- [29] E. J. Warrant, "Matched filtering and the ecology of vision in insects," in *The Ecology of Animal Senses*. Springer, 2016, pp. 143–167.
- [30] L. Freifeld, D. A. Clark, M. J. Schnitzer, M. A. Horowitz, and T. R. Clandinin, "Gabaergic lateral interactions tune the early stages of visual processing in drosophila," *Neuron*, vol. 78, no. 6, pp. 1075–1089, Jun. 2013.
- [31] 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, p. 427, Aug. 2014.
- [32] H. H. Yang, F. St-Pierre, X. Sun, X. Ding, M. Z. Lin, and T. R. Clandinin, "Subcellular imaging of voltage and calcium signals reveals neural processing in vivo," *Cell*, vol. 166, no. 1, pp. 245–257, Jun. 2016.
- [33] D. M. Bolzon, K. Nordström, and D. C. O'Carroll, "Local and large-range inhibition in feature detection," *J. Neurosci.*, vol. 29, no. 45, pp. 14 143–14 150, Nov. 2009.
- [34] J. C. Tuthill and B. G. Borghuis, "Four to fox trot: how visual motion is computed in the fly brain," *Neuron*, vol. 89, no. 4, pp. 677–680, Feb. 2016.
- [35] M.-J. Escobar, D. Pezo, and P. Orio, "Mathematical analysis and modeling of motion direction selectivity in the retina," *J. Phys. Paris*, vol. 107, no. 5, pp. 349–359, Nov. 2013.
- [36] A. D. Straw, "Vision egg: an open-source library for realtime visual stimulus generation," *Front. Neuroinf.*, vol. 2, p. 4, Nov. 2008.
- [37] N. Vogt and C. Desplan, "The first steps in drosophila motion detection," *Neuron*, vol. 56, no. 1, pp. 5–7, Oct. 2007.
- [38] L. Zheng, G. G. de Polavieja, V. Wolfram, M. H. Asyali, R. C. Hardie, and M. Jussola, "Feedback network controls photoreceptor output at the layer of first visual synapses in drosophila," *J. Gen. Physiol.*, vol. 127, no. 5, pp. 495–510, Apr. 2006.
- [39] T. Hwu, J. Isbell, N. Oros, and J. Krichmar, "A self-driving robot using deep convolutional neural networks on neuromorphic hardware," in *Proc. Int. Joint Conf. Neural Netw.*, 2017, pp. 635–641.
- [40] M. Beyeler, N. Oros, N. Dutt, and J. L. Krichmar, "A gpu-accelerated cortical neural network model for visually guided robot navigation," *Neural Netw.*, vol. 72, pp. 75–87, Dec. 2015.
- [41] N. A. Browning, S. Grossberg, and E. Mingolla, "Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance," *Neural Netw.*, vol. 22, no. 10, pp. 1383–1398, Dec. 2009.
- [42] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4996–5009, Sep. 2013.



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. He is currently pursuing the Ph.D. degree with the School of Computer Science, University of Lincoln, Lincoln, U.K.

His current research interests include small target detection models, artificial intelligence, and computational neuroscience.



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 is currently a Professor with the School of Mathematics and Information Science, Guangzhou University, Guangzhou, China. He was with Xi'an Jiaotong University and the City University of Hong Kong, Hongkong before 2018. His current research interests include nonlinear functional analysis and applications, machine learning theory, and sparse

information processing.



Shigang Yue (M'05–SM'17) received the B.Eng. degree from Qingdao University of Technology, Qingdao, China, in 1988, and the M.Sc. and Ph.D. degrees from the Beijing University of Technology, Beijing, China, in 1993 and 1996, respectively.

He is currently a Professor with the School of Computer Science, University of Lincoln, Lincoln, U.K. He was with the Beijing University of Technology, the City University of Hong Kong, Hong Kong, the University of Kaiserslautern, Kaiserslautern, Germany, University College London, London, U.K., Newcastle University, Newcastle upon Tyne, U.K., and the University of Cambridge, Cambridge, U.K., before 2007. His current research interests include artificial intelligence, computer vision, robotics, brains, and neuroscience.