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° to 3° of the visual region, with no response to larger bars (typically >10°) 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
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 direction 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.
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 filed 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 to photoreceptors. As depicted in Fig. 3, 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.

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,

\[
P(x, y, t) = \int \int I(u, v, t)G_{\sigma_1}(x - u, y - v)du dv
\]

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

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

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,

\[
L(x, y, t) = \int P(x, y, s)H(t - s)ds
\]

\[
H(t) = \Gamma_{n_1,\tau_1}(t) - \Gamma_{n_2,\tau_2}(t)
\]

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

\[
\Gamma_{n,\tau}(t) = (nt)^n \exp\left(-nt/\tau\right) \frac{\exp\left(-nt/\tau\right)}{(n-1)!\tau^{n+1}}.
\]

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,

\[
L_1(x, y, t) = \int \int L(u, v, s)W_1(x - u, y - v, t - s)du dv ds
\]
Here, $S_{Tm2}$ neurons respond strongly to luminance increments and of the $Tm1$ delayed relative to the $Tm2$ [32]. $S_{Mi1}$ and $Tm3$ respond selectively to luminance decrements, with the response of the $Mi1$ delayed relative to the $Tm3$ (see Fig. 1). The $Mi1$ and $Tm3$ respond selectively to $Tm1$, $Tm2$, $Tm3$ 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 for 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,

$$S_{Tm3}^{ON}(x, y, t) = S_{ON}^{ON}(x, y, t)$$
$$S_{Tm2}^{ON}(x, y, t) = S_{OFF}^{ON}(x, y, t)$$

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

$$S^{ON}(x, y, t) = [L_I(x, y, t)]^+$$
$$S^{OFF}(x, y, t) = [-L_I(x, y, t)]^-$$

where $S^{ON}(x, y, t)$ and $S^{OFF}(x, y, t)$ 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 $S_{Tm3}^{ON}(x, y, t)$ (or $S_{Tm2}^{ON}(x, y, t)$) with a Gamma kernel. That is,

$$S_{Tm3}^{Mi1}(n, \tau_N)(x, y, t) = \int S_{Tm3}^{ON}(x, y, s) \Gamma_{n, \tau_N}(t - s) ds$$
$$S_{Tm2}^{Mi1}(n, \tau_N)(x, y, t) = \int S_{Tm2}^{ON}(x, y, s) \Gamma_{n, \tau_N}(t - s) ds$$

where $S_{Tm3}^{Mi1}$ and $S_{Tm2}^{Mi1}$ represent the outputs of the $Mi1$ and $Tm1$, respectively. $n, \tau_N, \tau_F$ are orders of the Gamma kernels while $\tau_N, \tau_F$ 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 $S_{Tm3}^D$ and $S_{Tm2}^D$, are defined as,

$$S_{Tm3}^D(x, y, t) = \left[ \int \int S^{ON}(u, v, t)W_2(x - u, y - v) dudv \right]$$
$$S_{Tm2}^D(x, y, t) = \left[ \int \int S^{OFF}(u, v, t)W_2(x - u, y - v) dudv \right]$$

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

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

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

$$g(x, y) = G_{\sigma}(x, y) - e \cdot G_{\sigma}(x, y) - \rho$$

where $G_{\sigma}(x, y)$ is a Gaussian function and $e, \rho$ 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 $S_{Tm3}^D(x, y, t)$ and $S_{Tm2}^D(x, y, t)$ with a Gamma kernel. That is,

$$S_{Tm3}^{Mi1}(n, \tau_N)(x, y, t) = \int S_{Tm3}^D(x, y, s) \Gamma_{n, \tau_N}(t - s) ds$$
$$S_{Tm2}^{Mi1}(n, \tau_N)(x, y, t) = \int S_{Tm2}^D(x, y, s) \Gamma_{n, \tau_N}(t - s) ds$$

where $S_{Tm3}^{Mi1}$ and $S_{Tm2}^{Mi1}$ 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.
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,

\[
\tilde{D}(x, y, t) = \tilde{S}_{Tm1}^3(x, y, t) \times \tilde{S}_{D(n_3, \tau_3)}^1(x, y, t).
\]

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 \(\theta\), denoted by \(D(x, y, t; \theta)\), is defined as,

\[
D(x, y, t; \theta) = S_{Tm3}^1(x, y, t) \times \left\{ S_{D(n_3, \tau_3)}^1(x, y, t) + S_{Mi1}^1(x', y', t) \right\}
\times \tilde{S}_{D(n_3, \tau_3)}^1(x', y', t),
\]

where

\[
\begin{align*}
x' &= x + \alpha_1 \cos \theta \\
y' &= y + \alpha_1 \sin \theta
\end{align*}
\]

and \(\alpha_1\) 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., \(S_{D(n_3, \tau_3)}^1(x, y, t)\) and \(S_{Tm1}^1(x, y, t)\), the outputs of the Tm1 and Mi1 located at position \((x', y')\), i.e., \(S_{Tm1}^1(x', y', t)\) and \(S_{Mi1}^1(x', y', t)\) (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 \(\theta\). Thus, a series of correlation outputs \(D(x, y, t; \theta)\) with different preferred motion directions \(\theta\) can be defined. For a given direction \(\theta_0\), \(D(x, y, t; \theta_0)\) gives the strongest output to small target motion oriented along direction \(\theta_0\), 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,

\[
D_I(x, y, t; \theta) = \left[ \int \int D(u, v, t; \theta) W_2(x - u - v, y - v) du dv \right]^+.
\]

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 \(45^\circ\) apart by convolving \(D_I(x, y, t; \theta)\) with an inhibition kernel \(W_3(\theta)\). That is,

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

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

\[
W_3(\theta) = G_{\sigma_6}(\theta) - G_{\sigma_7}(\theta).
\]

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

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

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 (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 \(\theta\). That is,

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

where \(MD(t)\) denotes the motion direction of the small target at time \(t\), \((x, y)\) 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 two 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 a 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 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 to 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
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 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 \( \theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\} \). 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 \( \theta \) deviates from the small target motion direction.
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 \((n_4, \tau_4), (n_5, \tau_5)\) and \((\sigma_4, \sigma_5)\), 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. 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.
precisely, from Fig. 11(a) and (d), we can see that the increase of \((n_4, \tau_4)\) have not induced any significant changes of the Weber Contrast tuning curve and the height tuning curve. However, with the increase of \((n_4, \tau_4)\), as shown in Fig. 11(b) and (c), the peak velocity decreases while the peak width increases. In Fig. 12, the parameter \((n_5, \tau_5)\) has similar effect with \((n_4, \tau_4)\) 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 \((\sigma_4, \sigma_5)\). In this experiment, \((\sigma_4, \sigma_5)\) 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 \(\alpha_1\), \(\tau_4\) and \(\tau_5\) stand for the distance between position A and B, the peak velocity and the peak width, respectively. Then we have \(\Delta t_1 = \frac{\alpha_1}{v}, \Delta t_2 = \frac{\tau_4}{v}\) and \(\Delta t_3 = \frac{\tau_5}{v}\), respectively, where \(\alpha_1, v\) and \(w\) stand for the distance between position A and B, target width and velocity, respectively.

We further explain why \(\tau_4\) and \(\tau_5\) are positively correlated to \(\frac{\alpha_1}{v}\) and \(\frac{\tau_4}{v}\), respectively. 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 signals are: 1) ON signal of position A \((x, y)\), corresponding to \(S_{T \rightarrow M}^{(x, y),n_4}(x, y, t)\); 2) ON signal of position B \((x', y')\) with time delay order \(n_4\) and time delay length \(\tau_4\), corresponding to \(S_{D \rightarrow T}^{(x, y),n_4}(x', y', t)\); 3) OFF signal of position A \((x, y)\) with time delay order \(n_5\) and time delay length \(\tau_5\), corresponding to \(S_{D \rightarrow T}^{(x, y),n_5}(x, y, t)\); 4) OFF signal of position B \((x', y')\) with time delay order \(n_6\) and time delay length \(\tau_6\), corresponding to \(S_{D \rightarrow T}^{(x, y),n_6}(x', y', t)\). In the DSTMD, we set \(\tau_4\), \(\tau_5\) and \(\tau_6\) as \(\Delta t_1, \Delta t_2\) and \(\Delta t_1 + \Delta t_2\), respectively. Since \(\Delta t_1 = \frac{\alpha_1}{v}\) and \(\Delta t_2 = \frac{\tau_4}{v}\), we have \(\tau_4 = \alpha_1\) and \(\tau_5 = \frac{\tau_5}{v}\). That is, \(\tau_4\) and \(\tau_5\) are positively correlated to \(\frac{\alpha_1}{v}\) and \(\frac{\tau_4}{v}\), respectively.

In the third experiment, we demonstrate that the parameter \((\sigma_4, \sigma_5)\) 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 increase of \((\sigma_4, \sigma_5)\); 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 \(\sigma_4\) and \(\sigma_5\), where the higher \((\sigma_4, \sigma_5)\) 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 \(5 \times 5\) pixels, respectively. The coordinate of the small target at time \(t\) is \(500 - 250 \cdot \frac{4 + 300}{1000} \cdot 125 + 15 \cdot \sin(4\pi \frac{t + 300}{1000}), t \in [0, 1000]\) 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 \(\theta\) and the actual motion direction (shown in Fig. 17), the stronger DSTMD output tuned to this direction \(\theta\). 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.
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 \( \gamma \), if there is a position \((x_0, y_0)\), time \( t_0 \) and direction \( \theta_0 \) 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,

\[
D_R = \frac{\text{number of true detections}}{\text{number of actual targets}} \tag{32}
\]

\[
F_A = \frac{\text{number of false detections}}{\text{number of images}} \tag{33}
\]

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 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 pixels/s. A small target is moving against the cluttered background, and its coordinate at time \( t \) is \((500 - V_{x_T}t, 1000, 125 + 15 \cdot \sin(4\pi(t + 300) / 1000)), t \in [0, 1000]\) ms where \( V_{x_T} \) denotes the horizontal velocity. The receiver operating characteristic (ROC) curves of the three experiments with respect to target luminance, size and horizontal velocity \( V_{x_T} \) 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
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 $5 \times 5$ (or the velocity of 250) has higher detection rate compared to the target size of $3 \times 3$ and $8 \times 8$ (or the velocity of 200 and 350). This is because the ESTMD and DSTMD all exhibit selectivity. They show the strongest output to the target whose size (or velocity) equals $5 \times 5$ pixels (250 pixels/s), but weaker outputs to the object whose size (or velocity) is higher or lower than $5 \times 5$ pixels (250 pixels/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 pixels/s. A small target whose luminance and size are set as $0$ and $5 \times 5$ pixels, is moving against the cluttered backgrounds. The coordinate of the small target at time $t$ is $(500 - 250 \cdot t + 300, 125 + 15 \cdot \sin(4\pi \cdot t + 300))$, $t \in [0, 1000]$ 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.
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


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.