A Feedback Neural Network for Small Target Motion Detection in Cluttered Backgrounds

Hongxin Wang[1], Jigen Peng[2], and Shigang Yue[1]

[1] The Computational Intelligence Lab (CIL), School of Computer Science,
University of Lincoln, Lincoln, LN6 7TS, UK
`syue@lincoln.ac.uk`
[2] School of Mathematics and Information Science, Guangzhou University,
Guangzhou, 510006, China
`jgpeng@mail.xjtu.edu.cn`

Abstract. Small target motion detection is critical for insects to search for and track mates or prey which always appear as small dim speckles in the visual field. A class of specific neurons, called small target motion detectors (STMDs), has been characterized by exquisite sensitivity for small target motion. Understanding and analyzing visual pathway of STMD neurons are beneficial to design artificial visual systems for small target motion detection. Feedback loops have been widely identified in visual neural circuits and play an important role in target detection. However, if there exists a feedback loop in the STMD visual pathway or if a feedback loop could significantly improve the detection performance of STMD neurons, is unclear. In this paper, we propose a feedback neural network for small target motion detection against naturally cluttered backgrounds. In order to form a feedback loop, model output is temporally delayed and relayed to previous neural layer as feedback signal. Extensive experiments showed that the significant improvement of the proposed feedback neural network over the existing STMD-based models for small target motion detection.

Keywords: Small target motion detection · Feedback loop · Neural modeling · Naturally cluttered backgrounds

1 Introduction

In dynamic visual world, the observer (an animal) are more interested in moving objects, since they are more likely to be mates, predators or prey. Being able to detect moving objects in a distance and early could endow the observer with stronger competitiveness for survival. However, when an object is far away from the observer, it often appears as a small dim speckle whose size may vary from one pixel to a few pixels in the visual field. Detecting such small targets in visual cluttered backgrounds has been considered as a challenging problem for artificial visual systems. This is not only because shape, color and texture information of small targets cannot be used for motion detection, but also because the cluttered background, such as bushes, trees and/or rocks, always contains a great number

of small-target-like features (called background noise). Small target motion detection means detecting small moving targets, meanwhile discriminating them from background noise.

Insects exhibit exquisite sensitivity for small target motion [6] and can pursue small flying targets, such as mates or prey, with high capture rates [7]. As revealed in biological research [5, 6], the exquisite sensitivity is coming from a class of specific neurons in the insects' visual system, called small target motion detectors (STMDs). STMD neurons give peak responses to targets subtending $1-3°$ of the visual field, with no response to larger bars (typically $> 10°$) or to wide-field grating stimuli. The electrophysiological knowledge about STMD neurons and their afferent pathways is helpful for designing artificial visual systems for small target motion detection.

A few STMD-based models have been proposed for detecting small target motion in naturally cluttered backgrounds. Elementary small target motion detector (ESTMD) which was proposed by Wiederman *et. al.* [12], can detect the presence of small moving targets, but not the motion direction. To detect small moving targets and their motion directions, three directionally selective models have been proposed, including EMD-ESTMD [1, 11], ESTMD-EMD [1, 11] and directionally selective small target motion detector (DSTMD) [9]. Although these existing STMD-based models can detect small moving targets, their detection results often contain a great number of background noises. Further improvement is needed for filtering out background noises

Feedback loops exist extensively in animals' visual systems and can optimize motion estimation [3, 4]. Biological research reveals that feedback loops are able to simultaneously mediate the synthesis of motion representations and cancellation of distracting signals [3]. However, it is still unclear if a feedback loop exist in the visual pathway of STMD neurons or if a feedback loop can significantly improve detection performance of STMD neurons. In this paper, we investigate that if a feedback loop exists, can it improve detection performance of STMD neurons. To answer this question, we propose a feedback neural network (**feedback ESTMD**) based on the existing ESTMD model [12] for small target motion detection. In order to form a feedback loop, model output is firstly temporally delayed and then relayed to previous neural layer (medulla layer) as feedback signal. The feedback signal is added on the output of medulla layer for weakening responses to background noises. Systematic experiments demonstrate that the feedback loop can significantly improve detection performance of the existing STMD-based models.

The remainder of this paper is organized as follows. In section 2, the proposed feedback neural network is introduced in details. In section 3, experiments are carried out to test the performance of the proposed feedback neural network. Discussion is also given in this section. In section 4, we give conclusions and perspectives.
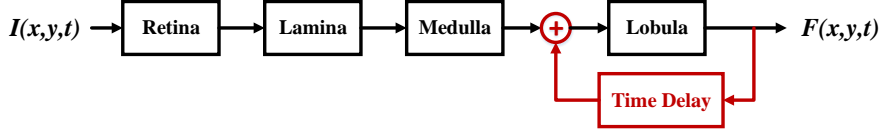
Fig. 1: Schematic illustration of the proposed feedback model.

2 Formulation of the Model

In this section, we elaborate on the proposed feedback model, called **Feedback ESTMD**. Its schematic illustration is shown in Fig. 1. As we can see from Fig. 1, $I(x, y, t)$ is the model input, denoting an image sequence where x, y and t are spatial and temporal field positions, respectively. Model input $I(x, y, t)$ is successively processed by four neural layers including retina, lamina, medulla and lobula. Through the process of four neural layers, we can obtain a model output $F(x, y, t)$. The output $F(x, y, t)$ is firstly temporally delayed and then relayed to medulla layer so as to form a feedback loop. The proposed feedback loop can weaken responses to background noises and significantly improve detection performance. In the following, functionalities of four neural layers and the feedback loop will be introduced in details.

2.1 Retina Layer

In the insect's visual system, retina layer contains a great number of ommatidia [10]. These ommatidia are able to receive luminance signals from the natural world and relay signals to downstream neurons for further process. The received luminance signal are always highly blurred, due to the extremely low resolution of ommatidia.

In the proposed feedback neural network, each ommatidium is modeled as a spatial Gaussian filter for simulating ommatidium's blur effect. Let $I(x, y, t) \in \mathbf{R}$ denote the input image sequence where x, y and t are spatial and temporal field positions. Then, the output of ommatidium with visual field centered at (x, y) denoted by $P(x, y, t)$ is defined by the following equation,

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

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

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

2.2 Lamina Layer

In the insect's visual system, lamina layer contains a great number of large monopolar cells (LMCs) [2]. LMCs receive signals from ommatidia and are able

to extract motion information from ommatidium output. To be more precise, LMCs show strong responses to brightness increments and decrements, i.e., luminance changes.

In the proposed feedback neural network, each LMC is modeled as a temporal high-pass filter extracting luminance changes, i.e., motion information, from ommatidium output $P(x, y, t)$. Let $L(x, y, t)$ denote the output of LMC located at (x, y). Then, $L(x, y, t)$ is defined by convolving ommatidium output $P(x, y, t)$ with a temporal high-pass convolution kernel $H(t)$. That is,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.3   Medulla Layer

In the insect's visual system, medulla layer contains a great number of medulla neurons, including Tm1, Tm2, Tm3 and Mi1 [2]. These four medulla neurons

receive signals from lamina layer and respond strongly to luminance changes. More precisely, Mi1 and Tm3 neurons respond selectively to brightness increments, with the response of Mi1 delayed relative to Tm3. Conversely, Tm1 and Tm2 respond selectively to brightness decrements, with the response of Tm1 delayed relative to Tm2.

Before modeling four medulla neurons, we firstly split LMC neural outputs $L_I(x, y, t)$ into positive and negative components denoted by $S^{ON}(x, y, t)$ and $S^{OFF}(x, y, t)$, respectively. That is,

$$S^{ON}(x, y, t) = [L_I(x, y, t)]^+ \tag{12}$$

$$S^{OFF}(x, y, t) = -[L_I(x, y, t)]^- \tag{13}$$

where $[x]^+, [x]^-$ denote $\max(x, 0)$ and $\min(x, 0)$, respectively. S^{ON} and S^{OFF} are also called ON and OFF signals, which are able to reflect brightness increments and decrements, respectively.

In the proposed feedback neural network, four medulla neurons including Tm1, Tm2, Tm3 and Mi1, not only receive feedforward signals from lamina layer, but also receive feedback signals from lobula layer (see Fig. 1). These two signals, i.e., feedforward and feedback signals, are added together to define the outputs of four medulla neurons.

Since Tm3 and Tm2 respond strongly to luminance increase and decrease, we firstly use $S^{ON}(x, y, t)$ and $S^{OFF}(x, y, t)$ to define the feedforward signals of Tm3 and Tm2, respectively. Let $F(x, y, t)$ denotes model output. Then, temporally delayed model output is used as the feedback signal. Finally, the output of Tm3 and Tm2 are defined by the addition of feedforward signal and feedback signal. That is,

$$S^{Tm3}(x, y, t) = \left[\iint S^{ON}(u, v, t) W_2(x - u, y - v) du dv \right]^+$$
$$\cdots + \boldsymbol{k} \cdot \int \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}) \boldsymbol{\Gamma}_{n_L, \tau_L}(\boldsymbol{t} - \boldsymbol{s}) d\boldsymbol{s} \tag{14}$$

$$S^{Tm2}(x, y, t) = \left[\iint S^{OFF}(u, v, t) W_2(x - u, y - v) du dv \right]^+$$
$$\cdots + \boldsymbol{k} \cdot \int \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}) \boldsymbol{\Gamma}_{n_L, \tau_L}(\boldsymbol{t} - \boldsymbol{s}) d\boldsymbol{s} \tag{15}$$

where S^{Tm3} and S^{Tm2} denote outputs of Tm3 and Tm2 neurons, respectively. k is a constant. n_L is order of Gamma kernel while τ_L is time constant. $W_2(x, y)$ is the second-order lateral inhibition kernel, defined as

$$W_2(x, y) = A[g(x, y)]^+ + B[g(x, y)]^- \tag{16}$$

where A, B are constant, and $g(x, y)$ is defined as

$$g(x, y) = G_{\sigma_4}(x, y) - e \cdot G_{\sigma_5}(x, y) - \rho \tag{17}$$

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

Since the neural response of Mi1 (or Tm1) is delayed relative to Tm3 (or Tm2), we define the feedforward signal of Mi1 (or Tm1) using temporally delayed output of Tm3 (or Tm2). Similarly, the feedback signal is added with the feedforward signal to define the output of Mi1 (or Tm1). That is,

$$S^{Mi1}(x, y, t) = \int \left[\iint S^{ON}(u, v, t) W_2(x - u, y - v) du dv \right]^+ \cdot \Gamma_{n_N, \tau_N}(t - s) ds$$
$$\cdots + k \cdot \int F(x, y, s) \Gamma_{n_L, \tau_L}(t - s) ds$$
(18)

$$S^{Tm1}(x, y, t) = \int \left[\iint S^{OFF}(u, v, t) W_2(x - u, y - v) du dv \right]^+ \cdot \Gamma_{n_F, \tau_F}(t - s) ds$$
$$\cdots + k \cdot \int F(x, y, s) \Gamma_{n_L, \tau_L}(t - s) ds$$
(19)

where S^{Mi1} and S^{Tm1} represent outputs of Mi1 and Tm1, respectively. n_N, n_F are orders of Gamma kernels while τ_N, τ_F are time constants.

2.4   Lobula Layer

In the insect's visual system, STMD neurons integrate signals from medulla neurons and respond selectively to small target motion.

In the proposed feedback neural network, the output of STMD neuron $F(x, y, t)$ with visual field centered at (x, y) is defined by multiplying Tm3 neural output $S^{Tm3}(x, y, t)$ with Tm1 neural output $S^{Tm1}(x, y, t)$. That is,

$$F(x, y, t) = S^{Tm3}(x, y, t) \cdot S^{Tm1}(x, y, t).$$
(20)

3   Results and Discussions

In this section, we test the ability of the proposed feedback neural network (Feedback ESTMD) for detecting small targets against cluttered backgrounds. The proposed neural network is tested on a set of image sequences produced by Vision Egg [8]. The video images are 500 (in horizontal) by 250 (in vertical) pixels and temporal sampling frequency is set as 1000 Hz.

Before performing experiments, we explain how to determine the location of a small moving target using model output $F(x, y, t)$. For a given detection threshold γ, if there is a position (x_0, y_0) and time t_0 which satisfy model output $F(x_0, y_0, t_0) > \gamma$, then we believe that a small target is detected at position
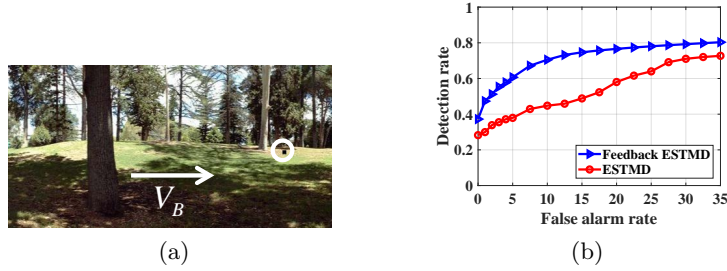
Fig. 2: (a) A representative frame of the input image sequence. The small target is highlighted by the white circle. Arrow V_B denote motion direction of the background. (b) The receiver operating characteristic (ROC) curve.

(x_0, y_0) and time t_0. Two metrics are defined to evaluate detection performance. That is,

$$D_R = \frac{\text{number of true detections}}{\text{number of actual targets}} \tag{21}$$

$$F_A = \frac{\text{number of false detections}}{\text{number of images}} \tag{22}$$

where D_R and F_A are 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 experiment, we use an image sequence which shows a small dark target moving against the naturally cluttered background, as model input. A representative frame is shown in Fig. 2(a). The background is moving from left to right and its velocity V_B is set as $V_B = 250$ (pixel/second). A small target is moving against the cluttered background and its coordinate at time t is set as $(500 - V_T \cdot \frac{t+300}{1000}, 125 + 15 \cdot \sin(4\pi \frac{t+300}{1000})), t \in [0, 1000]$ms where V_T denotes target velocity and is set as $V_T = 500$ (pixel/second). The luminance and size of the small target are set as 50 and 5×5 (pixel $\times$ pixel), respectively. The receiver operating characteristic (ROC) curve is presented in Fig. 2(b).

Fig. 2(b) is illustrating that the proposed feedback model (Feedback ESTMD) outperforms the existing model (ESTMD) at detecting small targets against naturally cluttered backgrounds. More precisely, for a given false alarm rate, feedback ESTMD has a higher detection rate than ESTMD. This also indicates that the feedback loop can improve detection performance of the existing STMD-based models.

We further test these two models under different parameters of the image sequence, including target luminance, target size, target velocity, background velocity and background motion direction. In order to compare detection performances, we fix false alarm rate F_A as 10 and illustrate detection rates of two models at this false alarm rate. The corresponding results are shown in Fig. 3.
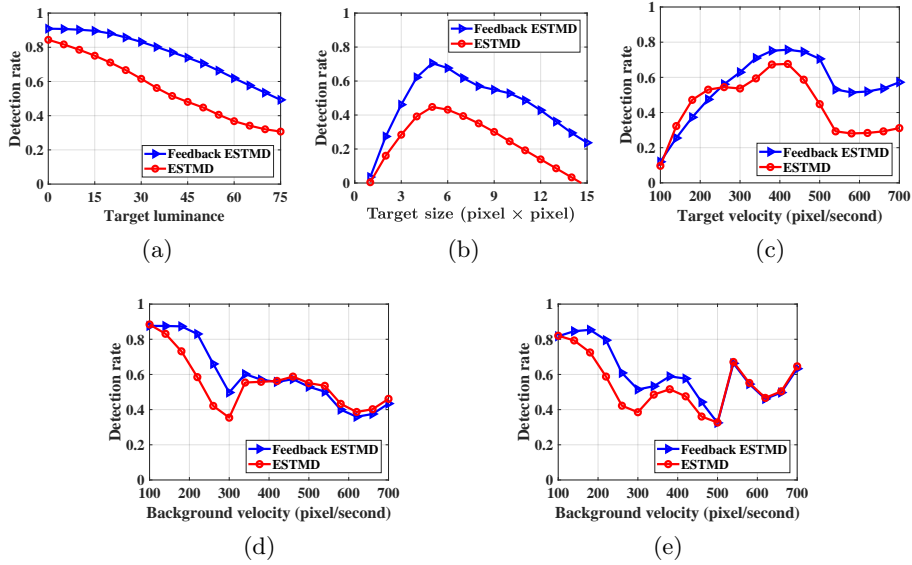
Fig. 3: Detection rates of the proposed feedback model (feedback ESTMD) and the existing model (ESTMD) at a fixed false alarm rate $F_A = 10$ when parameters of image sequences are changed. In each subplot, horizontal axis denotes the varying parameter while vertical axis denotes detection rate D_R. (a) Varying target luminance. (b) Varying target size. (c) Varying target velocity. (d) Varying background velocity when the target and the background are moving along the **opposite direction**. (e) Varying background velocity when the target and the background are moving along the **same direction**.

From Fig. 3(a) and (b), we can see that feedback ESTMD has a better detection performance than ESTMD under different target luminance and sizes. To be more precise, the detection rate of feedback ESTMD is much higher than that of ETMD when target luminance varies (see Fig. 3(a)). Similarly in Fig. 3(b), the detection rate of feedback ESTMD is higher than that of ETMD under different target sizes.

From Fig. 3(c), (d) and (e), we can find that detection performance of feedback ESTMD is dependent on velocity difference between the background and the small target. More precisely, as we can see from Fig. 3(c), when target velocity is larger than background velocity $V_B = 250$ (pixel/second), feedback ESTMD has higher detection rates than ESTMD. However, when target velocity is smaller than background velocity, detection rate of feedback ESTMD is slightly lower than that of ESTMD. Similar variation trend can be seen Fig. 3(d) and (e). To be more precise, no matter whether the background and the small target are moving along the same direction or not, the detection rate of feedback ESTMD is higher than that of ESTMD when background velocity is
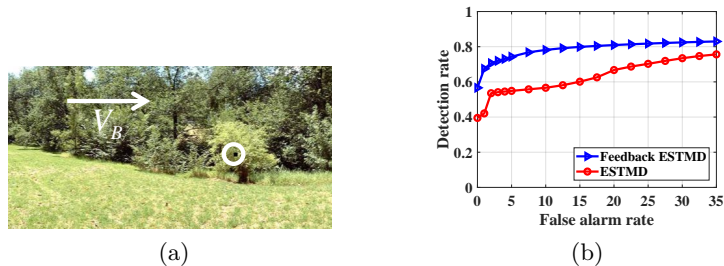
(a)

(b)

Fig. 4: (a) A representative frame of the input image sequence. The small target is highlighted by the white circle. Arrow V_B denote motion direction of the background. (b) The receiver operating characteristic (ROC) curves of feedback ESTMD and ESTMD.
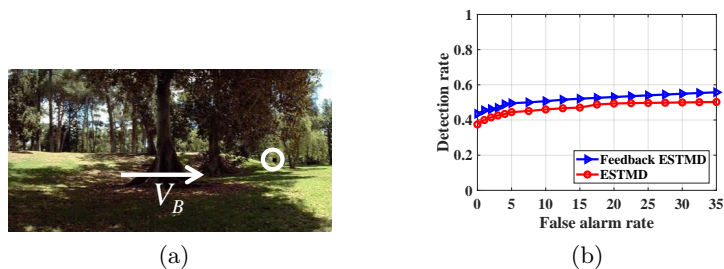

(a)

(b)

Fig. 5: (a) A representative frame of the input image sequence. The small target is highlighted by the white circle. Arrow V_B denote motion direction of the background. (b) The receiver operating characteristic (ROC) curves of feedback ESTMD and ESTMD.

smaller than target velocity $V_T = 500$ (pixel/second). When background velocity is larger than target velocity, detection rates of these two models show no significant difference.

In the second and third experiment, we test the proposed feedback model in different cluttered backgrounds. Two image sequences with different backgrounds are used as model input in these two experiments. Two representative frames are presented in Fig. 4(a) and Fig. 5(a), respectively. In these two image sequences, backgrounds are all moving from left to right and their velocities are set as 250 (pixel/second). A small target whose luminance, size are set as 50 and 5×5 (pixel $\times$ pixel), is moving against cluttered backgrounds. The coordinate of the small target at time t equals to $(500 - V_T \frac{t+300}{1000}, 125 + 15 * \sin(4\pi \frac{t+300}{1000})), t \in [0, 1000]$ where V_T is set as 500 (pixel/second).

From Fig. 4(b) and Fig. 5(b), we can see that feedback ESTMD has a better performance than ESTMD. For a given false alarm rate, the detection rate of feedback ESTMD is higher than that of ESTMD. This indicate that feedback ESTMD performs better than ESTMD in different cluttered backgrounds.

4   Conclusion

In this paper, we proposed a feedback neural network for small target detection against naturally cluttered backgrounds. In order to form a feedback loop, network output is temporally delayed and then relayed to middle neural layer as feedback signal. Feedback signal is added on outputs of middle neural layer for weakening responses to background noises. Systematic experiments showed that the proposed feedback neural network has a much better performance than the existing ESTMD model, if there is velocity difference between the background and the small target. In the future, we will further combine feedback loops with visual attention mechanisms for improving detection performances of models.

References

1. Bagheri, Z.M., Wiederman, S.D., Cazzolato, B.S., Grainger, S., OCarroll, D.C.: Performance of an insect-inspired target tracker in natural conditions. Bioinspiration & biomimetics 12(2), 025006 (Feb 2017)
2. Behnia, R., Clark, D.A., Carter, A.G., Clandinin, T.R., Desplan, C.: Processing properties of on and off pathways for drosophila motion detection. Nature 512(7515), 427 (2014)
3. Clarke, S.E., Maler, L.: Feedback synthesizes neural codes for motion. Current Biology 27(9), 1356–1361 (2017)
4. Kafaligonul, H., Breitmeyer, B.G., Öğmen, H.: Feedforward and feedback processes in vision. Frontiers in psychology 6, 279 (2015)
5. Nordström, K.: Neural specializations for small target detection in insects. Current opinion in neurobiology 22(2), 272–278 (2012)
6. Nordström, K., Barnett, P.D., O'Carroll, D.C.: Insect detection of small targets moving in visual clutter. PLoS biology 4(3), e54 (2006)
7. Olberg, R., Worthington, A., Venator, K.: Prey pursuit and interception in dragonflies. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology 186(2), 155–162 (Feb 2000)
8. Straw, A.D.: Vision egg: an open-source library for realtime visual stimulus generation. Frontiers in neuroinformatics 2, 4 (2008)
9. Wang, H., Peng, J., Yue, S.: A directionally selective small target motion detecting visual neural network in cluttered backgrounds. arXiv preprint arXiv:1801.06687 (2018)
10. Warrant, E.J.: Matched filtering and the ecology of vision in insects. In: The Ecology of Animal Senses, pp. 143–167. Springer (2016)
11. Wiederman, S.D., OCarroll, D.C.: Biologically inspired feature detection using cascaded correlations of off and on channels. Journal of Artificial Intelligence and Soft Computing Research 3(1), 5–14 (Dec 2013)
12. Wiederman, S.D., Shoemaker, P.A., O'Carroll, D.C.: A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS one 3(7), e2784 (2008)