Modeling Direction Selective Visual Neural Network with ON and OFF Pathways for Extracting Motion Cues from Cluttered Background

Qinbing Fu
School of Computer Science
University of Lincoln
Lincoln, LN6 7TS, UK
qifu@lincoln.ac.uk

Shigang Yue
School of Computer Science
University of Lincoln
Lincoln, LN6 7TS, UK
syue@lincoln.ac.uk

Abstract—The nature endows animals robust vision systems for extracting and recognizing different motion cues, detecting predators, chasing preys/mates in dynamic and cluttered environments. Direction selective neurons (DSNs), with preference to certain orientation visual stimulus, have been found in both vertebrates and invertebrates for decades. In this paper, with respect to recent biological research progress in motion-detecting circuitry, we propose a novel way to model DSNs for recognizing movements on four cardinal directions. It is based on an architecture of ON and OFF visual pathways underlies a theory of splitting motion signals into parallel channels, encoding brightness increments and decrements separately. To enhance the edge selectivity and speed response to moving objects, we put forth a bio-plausible spatial-temporal network structure with multiple connections of same polarity ON/OFF cells. Each pair-wised combination is filtered with dynamic delay depending on sampling distance. The proposed vision system was challenged against image streams from both synthetic and cluttered real physical scenarios. The results demonstrated three major contributions: first, the neural network fulfilled the characteristics of a postulated physiological map of conveying visual information through different neuropile layers; second, the DSNs model can extract useful directional motion cues from cluttered background robustly and timely, which hits at potential of quick implementation in vision-based micro mobile robots; moreover, it also represents better speed response compared to a state-of-the-art elementary motion detector.

I. INTRODUCTION

Extracting useful directional motion cues from dynamic and cluttered visual scenes is a critical ability for animals to locomote, avoid imminent collision, track moving targets. Several decades have witnessed a few milestones in motion-detecting circuit [1]–[4]. However, it is still pronounced elusive for the fundamental structure of how animals/humans retrieve meaningful visual information from the three dimensional world. Amongst lots of methodologies, the bio-plausible neural networks always provide us low-cost and efficient choices for processing visual signals.

There are overwhelming evidence that visual neurons respond to motion in a directionally selective way which have been found across animal kingdoms, the insects [1], [3] and also the vertebrates [2]. Those direction selective neurons (DSNs) in flies were first found in the lobula plate, which is one of the three stacks of neuropiles in the optic lobe, and they were also thought to be involved in visual flight control [4]. The lobula plate tangential cells (LPTCs) are known to be directionally selective with distinct preference: one system consists of the horizontal direction-sensitive (HS) cells, that depolarizing rigorously in response to horizontally moving stimulus in the preferred direction, while hyperpolarizing when stimulated by motions along the opposite direction; similarity for another group of the vertical direction-sensitive (VS) cells [5]. Interestingly, both HS and VS cells respond with a fixed preferred and non-preferred directions regardless of the color/contrast of the stimuli and background [5]. More specifically, HS cells prefer front-to-back (rightward) versus back-to-front directional motions, whilst VS cells prefer downward rather than upward directional stimuli [5].

In recent years, much progress have been made to demonstrate the motion detection mechanisms in insects [4]–[9], although the cellular implementations are largely unknown due to technical difficulties. Those works allow the following conclusions to be drawn: the motion detection starts with splitting the visual signals conveyed by the retinal into parallel ON and OFF pathways in the lamina area (first neuropile layer), encoding brightness increments (ON) and decrements (OFF) separately [4]–[7]. After that, there are specific cells synapsing ON units in the medulla (second neuropile layer) to form the ON pathway, and OFF units in the lobula (third layer) for the OFF pathway respectively [10]. The relayed visual signals are directionally selective with motion information calculated along each cardinal direction within ON/OFF pathway [10]. Finally, the outputs of different directional motion detectors are sorted and pooled in the lobula plate, i.e. the neurons regarding to the same directional tuning converge in an unanimous layer of lobula plate, jointly providing the input downstream to following circuits like the motion-driven systems [10]. In general, there are four sub-groups of LPTCs covering all cardinal directions, generating selective signals. Moreover, another important finding demonstrates visual signals in the final lobula plate which are delivered from the

dendrites of second and third neuropile layers are already directionally selective, which reveals the medulla/lobula layer conveys ON/OFF visual information separately and both layers might well be the place where same polarity signals from neighboring columns interact with each other [10].

In the following chapters, we will first review a few related works in Section II. The neural network description with algorithms will be illustrated in Section III. We then represent the systematic experiments and results in Section IV. Finally, we give a conclusion with future works in Section V.

II. RELATED WORK

Neurons compute the direction of motion corresponding to the well-known Hassenstein-Reichardt Correlation (HRC) model [1], which explains how nonlinear algorithm is mapped onto neuronal hardware and implemented by neural networks, also a milestone toward understanding how the brain computes based on sensory inputs. As a result, lots of associated works derive from the classic 'delay and correlate' model [3]–[5], [11]–[16]. With regard to research achievements in ON and OFF visual pathways, some hypothesis of either two or four channels for all pair-wise Reichardt detectors were proposed for explaining motion detection circuit in insects [6]. The existence of four-quadrant model with communications between both the same and opposite polarity channels mathematically conforms to the HRC model [1], [6]. Nevertheless, in accordance with the electrophysiological recordings from LPTCs, the two-quadrant detectors that processing only input combinations of the same sign (ON-ON and OFF-OFF), was demonstrated to exist in the motion detection circuitry rather than the four-quadrant detectors [6].

Another theory which was based on behavioral experiments, argued that either ON/OFF pathway conveys information about both positive and negative contrast changes to motion detection [7]. In this study, an architecture of combining six-quadrant detectors demonstrates ON/OFF cells preferentially process the motion of light/dark edges respectively, which also elicited the importance of edge selectivity in motion detection. To decide between these alternatives, a case study via genetically blocking either ON/OFF pathway gave strong evidence to the functional specialized two versus six quadrants signal flows [17]. In this work, we represented the interactions of same polarity signals via the putative two-quadrant Reichardt detectors. We also demonstrated the significance of edge selectivity when detecting motions embedded in cluttered environments.

Moreover, a profound implication is that a biological motion detection circuit cannot detect the true velocity of stimulus [12]. The Reichardt detector is elegant with intuitive computations; however, it still lacks of spatial and temporal comparison made between the internal components, such as the spatial separation of inputs, the temporal delays, the way of nonlinear interactions, each of which constrains the model performance [18]. In the speed tuning of previous models, detecting faster moving objects requires larger spatial span between sensitive neurons if the temporal delay is fixed; on the other hand, it needs shorter time delay when fixing the spatial span [18]. A
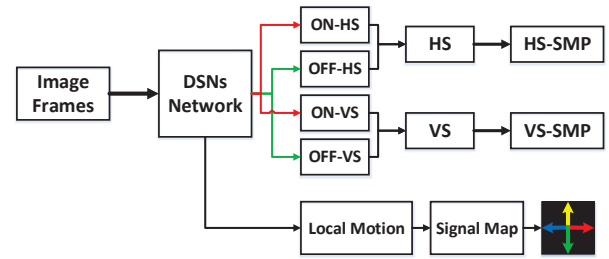

Fig. 1. DSNs neural network flowchart overview: inputs to model are gray-scaled successive image frames. After network processing, there are generally two kinds of outputs in parallel: first is local motion signal mapping which consists of direction information in four quadrants indicated by varied colors, as well as motion magnitude denoted by color densities; second are the sigmoid membrane potentials of both HS and VS cells, each of which is integrated by outputs from two sub-systems in ON and OFF pathways.

combination of multi-input Reichardt detectors was proposed to extract pattern-invariant velocities via spatial and temporal prefiltering of the correlated input [19]. In the proposed neural network, to enhance the speed response against moving objects, we investigated multi-connections of same polarity ON/OFF channels in spatial, along with dynamic temporal filtering for each pair-wised combination in sampling distance.

Furthermore, it is intriguing that the motion detection mechanisms varied in species. For example, even though the ON and OFF pathways have been identified in both flies and locusts, the characteristic array of DSNs found in the lobula plate of many flies are not found in locusts [20]. For locusts, the Lobula Giant Movement Detectors (LGMDs) have been pinpointed irreplaceable role of responding rigorously to moving objects in depth [21], [22]. In recent years, such looming-detecting neurons have been well studied and modeled for collisions prediction [9], [23]–[26]. Nevertheless, unlike DSNs, LGMDs cannot tell directional information of motions on four cardinal orientations. In addition, a synthetic visual neural network was proposed to construct DSNs via different directionally inhibitory convolutions [27], [28]. However, the major objectives of those works are also for collision recognition. Therefore, in the proposed modeling work, we were expected to build DSNs the specialists contributing to recognize useful motion cues in X-Y planes rather than movements in depth.

III. NETWORK ARCHITECTURE

In this section, we will present the DSNs model fully. The core concepts comprise an architecture of pair-wised symmetric Reichardt detectors contained in multi-connected ON-ON and OFF-OFF cells respectively, encoding brightness increments/decrements in ON/OFF pathway. Moreover, we put forward spatial and temporal filters to further enhance edge selectivity, as well as speed response to moving objects. A schematic model processing flowchart is illustrated in Fig. 1. The concrete neural network architecture is depicted in Fig. 2. In addition, a few related notation-abbreviations are listed in Table I.
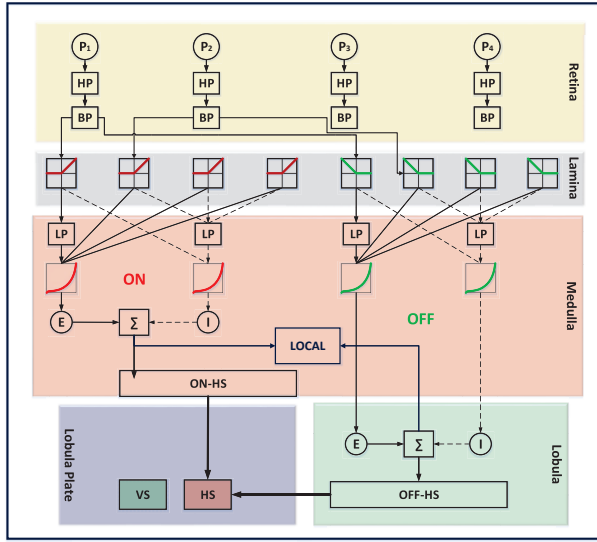
Fig. 2. The DSNs network schema throughout five labeled neuropile layers: To simplify, only the multi-connections of a single photoreceptor P_1 with P_2 until P_4 within HS system are depicted, wherein each P corresponds to an ON and OFF half-wave rectifier. HP and BP indicate high-pass and band-pass filters. LP, E and I denote low-pass filter, excitatory and inhibitory flows. $LOCAL$ is the local wide-field detector for motion signal map. In ON and OFF channels, the solid and dashed lines depict the nonlinear form of local excitation and inhibition respectively. Same directional flows from either pathway integrate in corresponded system of lobula plate. Similarity for the signal processing of other local pixels and also the units in VS system.

A. Retina Layer

In the retina, photoreceptors (P in Fig. 2) are in a two-dimensional matrix form to retrieve gray-level visual input. The number of photoreceptors corresponds to the resolution (C columns $\cdot$ R rows) of receptive field. A first-order high-pass filter (HP) follows with each photoreceptor, in order to get the differential image $L(t)$ between successive frames. We just take one cell as an example to express:

$$L(x, y, t) = P(x, y, t) - P(x, y, t - 1) \quad (1)$$

where $P(x,y,t)$ and $P(x,y,t-1)$ indicate the intensity on pixel (x,y) in current and previous frame respectively. After that, we apply a spatially band-pass filter (BP) represented by a 2-D form Difference of Gaussians (DoG), in order to enhance the motion edge selectivity, as well as removing redundant environmental noise so that maximizing information transmission. Such a mechanism is in accordance with the biological functions of large mono-polar cells, and has been suggested an appropriate filter characteristics prior to the site of motion detection [13], [16]. The center-positive Gaussian represents the excitatory receptive field L^e and the surround-negative one depicts the inhibitory field L^i:

$$L_{x,y}^e(t) = \int \int L(u, v, t) G_{\sigma_1}(x - u, y - v) d_u d_v,$$
$$L_{x,y}^i(t) = \int \int L(u, v, t) G_{\sigma_2}(x - u, y - v) d_u d_v \quad (2)$$

TABLE I
THE NOTATIONS OF DSNS MODEL

Notation: Abbreviation (Abbr), Full-name (Full)			
Abbr	Full	Abbr	Full
P	Photoreceptor	HS	Horizontal System
HP	Highpass Filter	VS	Vertical System
BP	Bandpass Filter	La	Lamina Layer
LP	Lowpass Filter	Me	Medulla Layer
E	Excitation	Lo	Lobula Layer
I	Inhibition	Lp	Lobula Plate

where the Gaussian function $G_\sigma(x, y)$ is defined as:

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \quad (3)$$

Since the differential image is convolved with two Gaussians of distinct covariances σ_1 and σ_2, the input can be thought of as including a set of overlapping fields. The center-surround field is created by having each point be the weighted average of the points surrounding it. And the weightings take a form of two Gaussian distributions. In DoG algorithm, the broader inhibitory Gaussian (with σ_2) is subtracted from the narrower excitatory one (with σ_1):

$$L_{x,y}^{DoG}(t) = L_{x,y}^e(t) - L_{x,y}^i(t) \quad (4)$$

B. Lamina Layer

Next in the lamina area, the first-order ON and OFF interneurons split signals relayed by the retina into parallel ON and OFF channels depending on brightness increments and decrements. The mechanism here is defined as the 'half-wave' rectifier (Fig. 2), filtering out or inverting negative input, along with brightness increments flowing into ON pathway, decrements of reverse-sign into OFF:

$$L_{x,y}^{ON}(t) = (L_{x,y}^{DoG}(t) + |L_{x,y}^{DoG}(t)|)/2,$$
$$L_{x,y}^{OFF}(t) = |(L_{x,y}^{DoG}(t) - |L_{x,y}^{DoG}(t)|)|/2 \quad (5)$$

Such a mechanism is also thought to be a central aspect before implementing arithmetic multiplication in the brain [7]. In addition, a Lipetz function with an exponent μ is applied to transform luminance analog value to membrane potential in a roughly logarithmic manner, which is conveyed to the processing of following neuropile layers:

$$La_{x,y}(t) = \frac{X(t)^\mu}{X(t)^\mu + X'(t)^\mu}. \quad (6)$$

where we let the $L_{x,y}^{ON}$ and $L_{x,y}^{OFF}$ be expressed by X. And the X' satisfies a mid-level parameter set as a first-order low-pass filtered version of X with time constant τ_1:

$$\frac{dX'(t)}{dt} = \frac{1}{\tau_1}(X(t) - X'(t)) \quad (7)$$
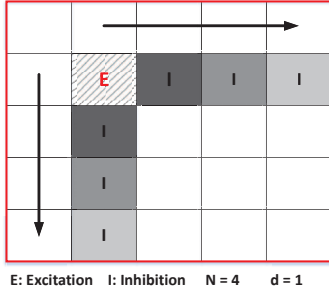
Fig. 3. The illustration of multi-connections for each local pixel in HS and VS with N denoting numbers of grouped sensitive cells (4 in this case), and d indicating the sampling distance increment (1 in this case). The excitations (E) are generated in the start cell, and the inhibitions (I) are formed in neighboring cells of both systems respectively. The different gray-scales of connected cells imply dynamic time delays with darker ones denoting longer delay and the increment of spatial distance leads to decrement of delay.

C. Medulla and Lobula Layers

Then ON and OFF cells relay visual information separately to neurons in the medulla and lobula layers (Fig. 2). Pertaining to the relevant biological research, the dendrites of neurons in these two layers are most likely the places where neighboring cells interact with each other in a nonlinear way to further generate directional selective signals to the lobula plate [10]. Compared to the state-of-the-art DSNs modeling works [27], [28], wherein the inhibitory connections for each local cell are on four/eight directions, we only map it with two directional connections as a result of the embedded symmetric Reichardt detectors. As shown in Fig. 3, for each local pixel, we link it with defined number of adjacent cells only in feed forward directions: front-to-back for HS and downward for VS. We also set a spatial span corresponding the increment of sampling distance within every connection.

The excitations in direction-sensitive cells were put forward to symmetrically arrange around dendrites [29]. When stimulated by motion in preferred directions, inhibitory signals lag behind excitatory flows; otherwise, inhibitions arrive before excitations in non-preferred directions [29]. It fully satisfies the structure of symmetric Reichardt detectors, which give implications of mathematically bridging it with each pair-wised combination in either pathway. The temporal delays still reconcile with the first-order low-pass filters (LP), however, with dynamic pair-wised time constant τ_s. For each local cell, we put forth the longest time span in the nearest pair in sampling distance, and gradually reduce the delay as growing of the spatial span along both directional multi-connections (Fig. 3). Let $La_{x,y}$ be X and the delayed information $D_{x,y}$ be Y, we could deduce the following equation:

$$\frac{dY(t)}{dt} = \frac{1}{\tau_s}(X(t) - Y(t)) \qquad (8)$$

The signals then undergo the multiplications to form excitation (E) in the start cell and inhibitions (I) in all connected cells (Fig. 3). Those two kinds of signals interact at each local pixel as shown in Eq. 9 for ON channels of HS and in Eq. 10 for

OFF channels of HS. Similarity for the calculations in both pathways of VS system which are not stated here.

$$E_{x,y,HS}^{ON}(t) = \sum_{i=d}^{d \cdot N} D_{x,y}^{ON}(t) \cdot La_{x+i,y}^{ON}(t),$$
$$I_{x,y,HS}^{ON}(t) = \sum_{i=d}^{d \cdot N} D_{x+i,y}^{ON}(t) \cdot La_{x,y}^{ON}(t) \qquad (9)$$

$$E_{x,y,HS}^{OFF}(t) = \sum_{i=d}^{d \cdot N} D_{x,y}^{OFF}(t) \cdot La_{x+i,y}^{OFF}(t),$$
$$I_{x,y,HS}^{OFF}(t) = \sum_{i=d}^{d \cdot N} D_{x+i,y}^{OFF}(t) \cdot La_{x,y}^{OFF}(t) \qquad (10)$$

Then for each local cell, the excitation is linearly subtracted by inhibition. A gain factor w_i is embedded in every inhibitory flow, to form a partially balanced model [6]:

$$Me_{x,y,HS}(t) = E_{x,y,HS}^{ON}(t) - w_i \cdot I_{x,y,HS}^{ON}(t),$$
$$Lo_{x,y,HS}(t) = E_{x,y,HS}^{OFF}(t) - w_i \cdot I_{x,y,HS}^{OFF}(t) \qquad (11)$$

As depicted in Fig. 2, there are also local wide-field motion detectors concurrently gathering directional information passed in both polarity channels, wherein signals merge both linearly and multiplicatively as proposed in [8]. These combinations of term coefficients $\theta_1, \theta_2, \theta_3$ allow us to adjust different balances of contributions between separated ON and OFF channels. Let $Me_{x,y,HS}(t)$ be X, $Lo_{x,y,HS}(t)$ be Y, we could define the equation below:

$$W_{x,y,HS}(t) = \theta_1 \cdot X + \theta_2 \cdot Y + \theta_3 \cdot X \cdot Y \qquad (12)$$

Therefore, we can calculate the local motion direction Dir and magnitude M for mapping the signal with distinctive directional motion information (Fig. 1). Let $W_{x,y,HS}(t)$, $W_{x,y,VS}(t)$ be X and Y, we could infer the following equations:

$$Dir_{x,y}(t) = \arctan 2(Y, X) \cdot 180/\pi,$$
$$M_{x,y}(t) = \sqrt{X^2 + Y^2} \qquad (13)$$

where $\arctan 2$ satisfies the 'arctangent' function with two arguments in order to return the radian of computed angle in appropriate quadrant.

D. Lobula Plate

Four sub-groups of directional signals conveyed by the neurons in medulla of ON pathway and lobula of OFF pathway constitute the final HS and VS motion detection systems as illustrated in both Fig. 1 and 2. The direction-specific visual signals congregate in the corresponded layer of lobula plate (Lp), towards further motion control system. With respect to the symmetric structure, for both HS and VS outputs, positive membrane potential will be invoked by stimuli in preferred

TABLE II
THE PARAMETERS SETTING OF DSNs MODEL

Parameter: Name, Value (Val)					
Name	Val	Name	Val	Name	Val
N	$4 \sim 8$	σ_1	$0.5 \cdot N$	τ_1	750ms
d	$1 \sim 4$	σ_2	$0.9 \cdot N$	τ_s	$1 \sim 200$ms
C	$320, 432$	Δ_C	0.5	θ_1	$0, 1, 2$
R	240	K	0.01	θ_2	$0, 1, 2$
w_i	0.9	μ	0.7	θ_3	$0, 1, 2, 6$

(a) dark approaching-receding (b) light approaching-receding

Fig. 4. The outputs of DSNs model against dark (a)/light (b) objects approaching-receding embedded in bright/dark background. The red line denotes the image-size change in the view. X and Y-axis indicate time course in frames and potential (SMP) level. Both HS and VS SMPs are depicted.

directions, yet negative in non-preferred ones. The two sub-systems of HS illustrated in Fig. 1 and 2 are integrated by the following equations:

$$Lp_{HS}^{ON}(t) = \sum_1^C \sum_1^R Me_{x,y,HS}(t)$$
$$Lp_{HS}^{OFF}(t) = \sum_1^C \sum_1^R Lo_{x,y,HS}(t)$$
(14)

Similarity for VS which is not stated here. Before pooling in final HS and VS of lobula plate, the membrane potentials in the four directional layers go through a sigmoid transformation which mimics the activation function of artificial neurons to generate the sigmoid membrane potential (SMP) [9]:

$$f(x) = (1 + e^{-|x| \cdot (C \cdot R \cdot K)^{-1}})^{-1} - \Delta_C$$
(15)

where a coefficient K shapes the nonlinear function curve. And a small real number Δ_C ranges the output within $[0, 0.5)$ for positive input; otherwise, $f(x) = -f(x)$ in the scope of $(-0.5, 0]$ for negative input. It is important to note that the directionally selective signals are eventually integrated to form two flows in HS and VS systems (Fig. 1 and 2) in parallel as the network outputs, each of which is ranged in $(-1, 1)$:

$$HS(t) = Lp_{HS}^{ON}(t) + Lp_{HS}^{OFF}(t),$$
$$VS(t) = Lp_{VS}^{ON}(t) + Lp_{VS}^{OFF}(t)$$
(16)

It is worth emphasizing here, the proposed visual neural network only involves low-level image processing, wherein those computational expensive methodologies such as the target classification, scene analysis are not applied at all, which hits at potential of quick implementation in vision-based platforms, like the robots.

IV. EXPERIMENTS AND ANALYSIS

In this section, we will present the systematic experiments to examine the performance of DSNs model. In general, we set up two kinds of off-line tests that the vision system was challenged against image streams from synthetic and cluttered real-physical scenes respectively. In the former kind of synthetic tests, to investigate the functional specialized ON and OFF pathways, we sieved either ON/OFF pathway respectively. In the later kind of trials under stimuli embedded in cluttered background, in order to evaluate the enhanced speed response of this neural network, we put up comparative
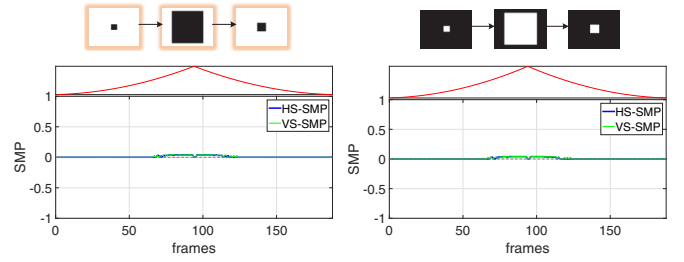
experiments for comparing it with a state-of-the-art computational model which is also based on Reichardt detectors with signal processing only in a single channel [14].

A. Parameter and Software Setting

First of all, the network parameters were decided empirically with consideration of functionalities. There are no parameters training and learning methods currently being involved in. Table II lists the parameters setting up. It is necessary to note here the sampling distance N corresponds to the calculation of covariances σ_1, σ_2 in DoG of band-pass filter. The dynamic time τ_s in temporal low-pass filtering varied within a wide range in milliseconds.

The computer used was a laptop with two 2.30 GHz CPUs. The DSNs model was built in Visual Studio 2013 (Microsoft Corporation). Experimental data plotting were implemented in Matlab 2015b (The MathWorks, Inc. Natick, USA). The input visual stimuli were all converted to gray-scale format. The resolutions of image streams were $320 \cdot 240$ and $432 \cdot 240$ for simulated and real-physical scenes respectively.

B. Challenged against Synthetic Stimuli

The computer-simulated stimuli include dark/bright objects approaching-receding (Fig. 4), translating (Fig. 5) and elongating-shortening (Fig. 6) against light/dark backgrounds respectively. There was no background noise in such synthetic scenarios. We also investigated the speed response of the DSNs model challenged by X-Y planes movements with varied constant-speed levels (Fig. 7).

First, when facing with either dark/bright objects moving in depth, the DSNs model represents very weak response as illustrated in Fig. 4. Although we put forth a partial-balanced structure in ON and OFF pathways that motion in preferred versus non-preferred directions will evoke stronger response when combining with two kinds of signals (excitation and inhibition) as presented in last section, the objects' edges almost simultaneously converging or diverging along four cardinal directions during moving in depth, which tends to cancel out most HS and VS responses. Therefore, the proposed neural network is not suitable for perceiving collision. However, it is in complementary with the looming-detecting neuron models
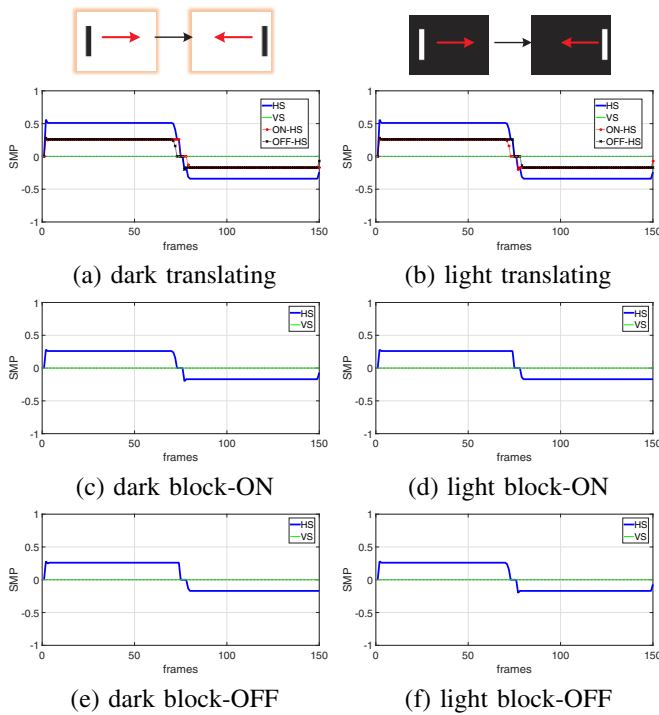
835

Fig. 5. The outputs of DSNs model against dark/light translation stimuli on two horizontal directions at a constant speed: (a), (b) VS/HS (with two sub-HS) SMPs (c), (d) VS/HS readouts when blocking ON pathway (e), (f) VS/HS outputs when sieving OFF pathway. Similarity for other notations.
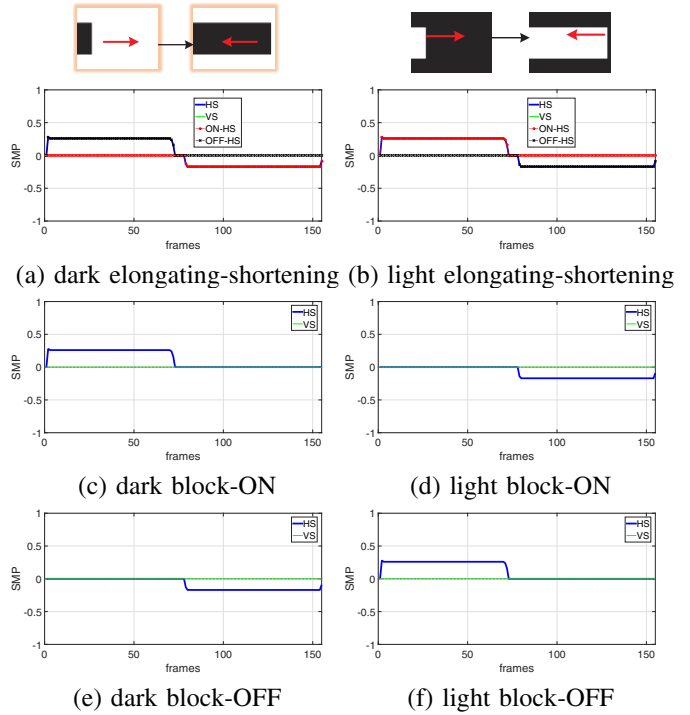

Fig. 6. The outputs of DSNs model against dark/light elongating-shortening stimuli with a single moving edge at a constant speed: (a), (b) VS/HS (with two sub-HS) outputs (c), (d) HS/VS readouts when sieving ON pathway (e), (f) HS/VS outputs when blocking OFF pathway. Similarity for other notations.

which can not extract directional motion cues from movements on four cardinal orientations [9], [23], [24], [26].

Second, on the aspect of translation on directions of front-to-back/back-to-front (Fig. 5), for the dark moving object, the leading edge invokes offset response whilst the trailing edge meanwhile brings about onset response. Therefore, as shown in Fig. 5, the two HS sub-systems in ON and OFF pathways both elicit continuous positive membrane potentials for motions in preferred direction (front-to-back), negative ones for the non-preferred directional movements (back-to-front). In addition, it represents more vigorous responses to motions in preferred direction which demonstrates an imbalance of the symmetric pair-wised Reichardt detectors in both pathways. On the contrary, VS system keeps quiet throughout each intact process. Similarity for the light translating stimuli, whereas the leading/trailing edges activate ON/OFF cells respectively. Intriguingly, when blocking either pathway, the SMPs are curtailed to the half-level: sieving ON pathway abolishes response elicited by ON cells leaving that of OFF cells unchanged; otherwise sieving OFF pathway gives rise to only onset response. The results adequately reconcile with related biological research [17].

Moreover, the elongating-shortening stimuli represent only a single edge movements: dark-elongating and light-shortening lead to light-to-dark luminance change of offset response; the opposite directional movements correspond onset events. As illustrated in Fig. 6, compared to the outputs against translation, only one HS sub-system contributes to HS output: ON-
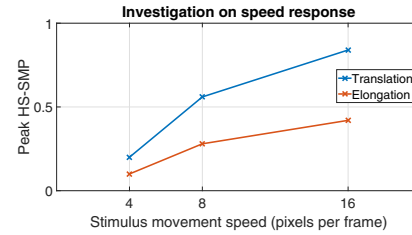

Fig. 7. Peak HS SMPs against translation/elongation stimuli along preferred direction (front-to-back) at three constant-speed levels.

HS cells elicit positive SMP for light-elongating, negative for dark-shortening; OFF-HS cells elicit positive SMP when dark-elongating, negative when light-shortening. Blocking either ON/OFF pathway remains only the opposite polarity channel (OFF/ON) effective. Furthermore, we investigated the speed response of the DSNs model against translation and elongation stimuli with different constant speeds. It is clear to see from Fig. 7, the peak SMP climbs up along with objects moving faster. This suggests the neural network is sensitive to the speed of stimulus.

To give a brief summary for synthetic off-line tests, the proposed neural network performs robustly in recognizing directional motion cues, and perceiving moving objects in four cardinal directions rather than looming in depth. To further pinpoint its effectiveness of motion detection in real world, we challenged it by more complex movements embedded in cluttered real physical scenarios.

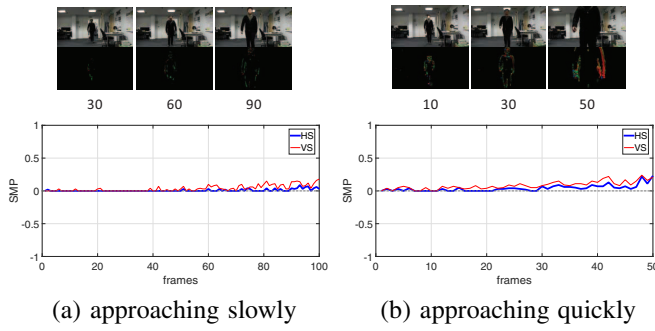(a) approaching slowly   (b) approaching quickly

Fig. 8. The outputs of DSNs model against a person approaching embedded in cluttered background from a fixed start position at two speed levels. A few frame-labeled snapshots with corresponded motion signal maps are shown at top. X-axis denotes the time sequence. Both HS and VS outputs are depicted.

C. Challenged against Cluttered Background

In comparison with former kind of tests, there was background noise for real physical scenarios. We used a fixed camera to record each motion process. First, the stimuli consist of a person approaching (Fig. 8) and crossing (Fig. 9) against cluttered background. As illustrated in Fig. 8, the DSNs model represents very weak responses of both HS and VS during either slow/fast approaching, even after the visual stimulus getting close. In addition, the local motion signal maps demonstrate the edge selectivity of DSNs model: the background noise was rigorously reduced leaving most meaningful motion cues. When challenged by crossing stimuli, Fig. 9 demonstrates that distinctive directional motion cues have been extracted from the cluttered background via both motion signal maps and SMP mapping: the DSNs model represents positive/negative SMPs of HS with its two ON/OFF sub-systems for motions in preferred/non-preferred directions. In this case, the VS response stays at much lower level. Moreover, motions in either direction with higher velocity give rise to more significant ramp up/down of SMPs (Fig. 9). The model outputs are also stronger in preferred over non-preferred direction like the results of former tests (Fig. 5 and 6). The motion signal maps also verify the successful recognition of directional motion cues with edge selectivity in even cluttered background. Then, we repeated approaching-receding and translation stimuli using a dark ball. As illustrated in Fig. 10, the outputs of DSNs model approve its robustness of extracting motion cues on cardinal directions, whereas it is rigorously inhibited by movements in depth.

As mentioned above, we also set up systematic translation experiments to examine the speed response of the DSNs model compared to a HRC-based EMDs model [14]. The visual stimuli were same as the ball translating front-to-back as shown in Fig. 10, but with three approximately constant speed levels respectively. The statistical results in Fig. 11 reveal a better speed response of the proposed DSNs model: the stimuli at faster angular speed leads to higher peak response though weaker increasing after 10 rad/s; on the other hand, the HRC-based model represents a decreasing peak outputs. The enhanced speed response is achieved via the multi-connections

(a) front-to-back slow   (b) front-to-back fast

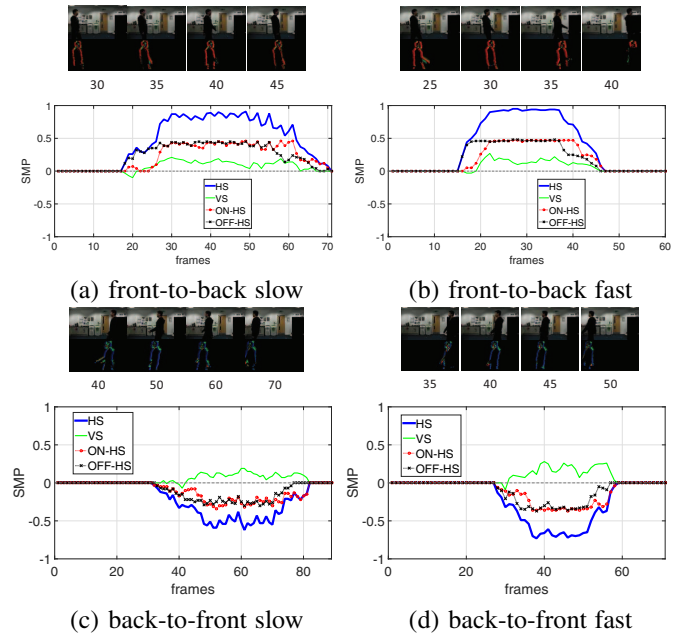(c) back-to-front slow   (d) back-to-front fast

Fig. 9. The outputs of DSNs model against a person crossing the view within cluttered background on two directions respectively at two speed levels. VS/HS (with two sub-HS) readouts are depicted. Similarity for other notations.

(a) ball approaching   (b) ball receding
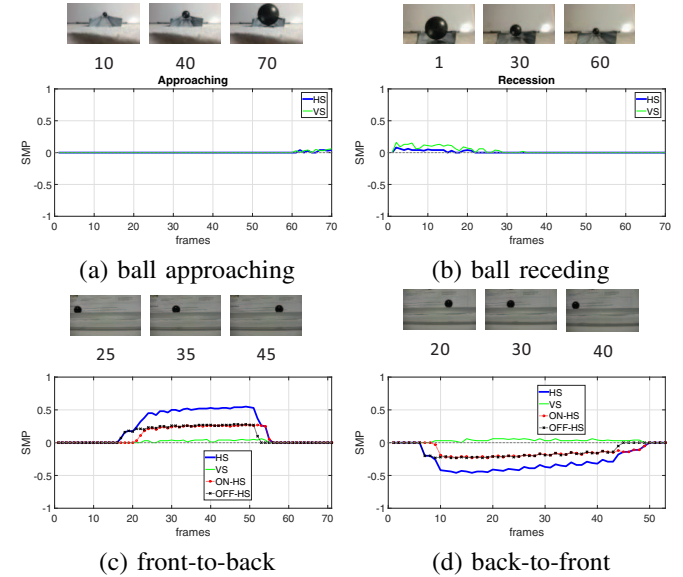

(c) front-to-back   (d) back-to-front

Fig. 10. The outputs of DSNs model against a dark ball approaching (a), receding (b), translation front-to-back (c), and back-to-front (d).

of same polarity cells in spatial with dynamic delays in temporal filtering.

V. CONCLUSION AND FUTURE WORK

To summarize, in this paper, we have proposed a visual neural network for modeling direction selective neurons with ON and OFF pathways, which fulfills the characteristics of a putative physiological motion detection circuit. The core concept of this model is that the visual signals are split into two parallel ON and OFF pathways with interactions of only the
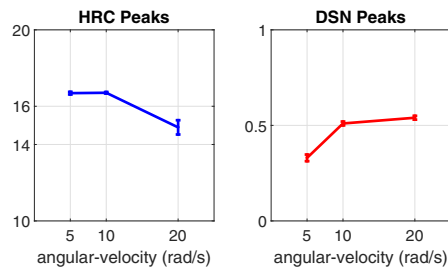
Fig. 11. Statistical results of two vision systems accounting for translation stimuli at three varied angular velocities, each of which repeated three times respectively. X-axis indicates translating angular velocity in radian per second. The left sub-figure represents logarithmic peak responses of a HRC-based model [14]. The right one depicts peak-SMPs of the proposed DSNs model.

same polarity channels. The visual information are processed separately within direction-specific pathways throughout the whole circuitry. Through our systematic experiments, we have demonstrated its spatial and temporal characteristics of extracting useful motion cues from cluttered environments via a novel structure of multi-connected ON-ON/OFF-OFF afferent cells with dynamic delays. The experimental results have illustrated how the four-cardinal directional information are mapped by the membrane potential, and approved the proposed neural network can perceive motions with edge selectivity and enhanced speed response. Since the bio-plausible DSNs model fully reconciles with related researches in signals tuning/processing in animals' vision system, it may bring benefit toward understanding the very complex motion detection mechanism. Moreover, compared to a few alternative vision systems in motion-flows detection such as the convolutional networks, the DSNs model depicts a feed forward network structure and only involves low-level image processing methods which may also shed light on efficient implementation in vision-based robots.

What does the future hold? An interesting question concerns on employing this neural network in object tracking. And since its functionality is in complementary with the looming-detecting neuron models, we are also prospective of building multiple pathways dealing with more complicated visual tasks.

ACKNOWLEDGMENT

REFERENCES

[1] V. Hassenstein and W. Reichardt, "System theoretical analysis of time, sequence and sign analysis of the motion perception of the snout-beetle chlorophanus," *Z Naturforsch*, vol. 11b, pp. 513–524, 1956.

[2] H. Barlow and W. Levick, "The mechanism of directionally selective units in rabbit's retina," *The journal of physiology*, vol. 178(3), pp. 477–504, 1965.

[3] E. Buchner, "Elementary movement detectors in an insect visual system," *Biol. Cybern*, vol. 24, p. 85101, 1976.

[4] A. Borst and T. Euler, "Seeing things in motion: models, circuits, and mechanisms," *Neuron*, vol. 71, no. 6, pp. 974–994, 2011.

[5] M. Joesch, B. Schnell, S. V. Raghu, D. F. Reiff, and A. Borst, "On and off pathways in drosophila motion vision," *Nature*, vol. 468, no. 7321, pp. 300–304, 2010.

[6] 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, 2011.

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

[8] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, "Correlation between off and on channels underlies dark target selectivity in an insect visual system," *J Neurosci*, vol. 33, no. 32, pp. 13 225–13 232, 2013.

[9] S. Yue and F. C. Rind, "Collision detection in complex dynamic scenes using a lgmd based visual neural network with feature enhancement," *IEEE Trans. Neural Netw*, vol. 17, no. 3, p. 705716, 2006a.

[10] M. S. Maisak, J. Haag, G. Ammer, E. Serbe, M. Meier, A. Leonhardt, T. Schilling, A. Bahl, G. M. Rubin, A. Nern, B. J. Dickson, D. F. Reiff, E. Hopp, and A. Borst, "A directional tuning map of drosophila elementary motion detectors," *Nature*, vol. 500, no. 7461, pp. 212–216, 2013.

[11] F. Gabbiani and P. W. Jones, "A genetic push to understand motion detection," *Neuron*, vol. 70, no. 6, pp. 1023–1025, 2011.

[12] M. Frye, "Elementary motion detectors," *Curr Biol*, vol. 25, no. 6, pp. R215–217, 2015.

[13] A. Borst and M. Egelhaaf, "Principles of visual motion detection," *Trends Neurosci*, vol. 12, no. 8, pp. 297–306, 1989.

[14] F. Iida and D. Lambrinos, "Navigation in an autonomous flying robot by using a biologically inspired visual odometer," *Proc. SPIE 4196, Sensor Fusion and Decentralized Control in Robotic Systems III*, vol. 4196, pp. 86–97, 2000.

[15] J. M. Zanker and J. Zeil, "Movement-induced motion signal distributions in outdoor scenes," *Network: Computation in Neural Systems*, vol. 16, no. 4, pp. 357–376, 2005.

[16] A. Pallus and L. J. Fleishman, "A two-dimensional visual motion detector based on biological principles," https://muse.union.edu/visualmotion/main-article/, accessed: 2014-07-01.

[17] M. Joesch, F. Weber, H. Eichner, and A. Borst, "Functional specialization of parallel motion detection circuits in the fly," *J Neurosci*, vol. 33, no. 3, pp. 902–905, 2013.

[18] M. J. Zanker, V. M. Srinivasan, and M. Egelhaaf, "Speed tuning in elementary motion detectors of the correlation type," *Biological Cybernetics*, vol. 80, no. 2, pp. 109–116, 1999.

[19] H. P. Snippe and J. J. Koenderink, "Extraction of optical velocity by use of multi-input reichardt detectors," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1222–1236, 1994.

[20] S. Wernitznig, F. C. Rind, P. Polt, A. Zankel, E. Pritz, D. Kolb, E. Bock, and G. Leitinger, "Synaptic connections of first-stage visual neurons in the locust schistocerca gregaria extend evolution of tetrad synapses back 200 million years," *J Comp Neurol*, vol. 523, no. 2, pp. 298–312, 2015.

[21] F. C. Rind and B. D. I., "Neural network based on the input organization of an identified neurone signaling impending collision," *J Neurophysiol*, vol. 75, pp. 967–985, 1996.

[22] 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, pp. 203–214, 1997.

[23] Q. Fu, C. Hu, and S. Yue, "Bio-inspired collision detector with enhanced selectivity for ground robotic vision system," in *British Machine Vision Conference (BMVC)*. York, UK, 2016, Conference Proceedings.

[24] Q. Fu and S. Yue, "Modelling lgmd2 visual neuron system," in *Machine Learning for Signal Processing (MLSP)*. 2015 IEEE 25th International Workshop, Boston, MA, 2015, Conference Proceedings.

[25] C. Hu, F. Arvin, C. Xiong, and S. Yue, "A bio-inspired embedded vision system for autonomous micro-robots: the lgmd case," *IEEE Transactions on Cognitive and Developmental Systems*, 2016.

[26] S. Yue and F. C. Rind, "Redundant neural vision systemscompeting for collision recognition roles," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 2, pp. 173–186, 2013.

[27] ——, "A synthetic vision system using directional selective motion detectors for collision recognition," *Artificial Life*, vol. 13, no. 2, pp. 93–122, 2007.

[28] ——, "Postsynaptic organization of directional selective visual neural networks for collision detection," *Neurocomput*, vol. 103, pp. 50–62, 2013.

[29] W. F. Fried SI., Mnch TA., "Mechanisms and circuitry underlying directional selectivity in the retina," *Nature*, vol. 420, pp. 411–414, 2002.