ORIGINAL ARTICLE

Bio-inspired visual neural network on spatio-temporal depth rotation perception

Bin Hu[1] · Zhuhong Zhang[2] ⓘ

Abstract

In primates' cerebral cortex, depth rotation sensitive (DRS) neurons have the property of preferential selectivity for depth rotation motion, whereas such a property is rarely adopted to create computational models for depth rotation motion detection. To fill this gap, a novel feedforward visual neural network is developed to execute depth rotation object detection, based on the recent neurophysiologic achievements on the mammalian vision system. The proposed neural network consists of two parts, i.e., presynaptic and postsynaptic neural networks. The former comprises multiple lateral inhibition neural sub-networks for the capture of visual motion information, and the latter extracts the cues of translational and depth motion and later, synthesizes such clues to perceive the process of depth rotation of an object. Experimentally, the neural network is sufficiently examined by different types of depth rotation under multiple conditions and settings. Numerical experiments show that not only it can effectively detect the spatio-temporal energy change of depth rotation of a moving object, but also its output excitation curve is a quasi-sinusoidal one, which is compatible with the hypothesis suggested by Johansson and Jansson in projective geometry. This research is a critical step toward the construction of artificial vision system for depth rotation object recognition.

Keywords Depth rotation sensitive neuron · Quasi-sinusoidal waveform · Lateral inhibition · Depth rotation · Visual motion perception

1 Introduction

Visual motion perception is a challenging topic in computer vision. Visual scene is a wealth of information which depicts the motion of an object in external environments, by which motion cues can be extracted by visual information processing in order to design intelligent vision systems [1, 2]. However, it is still difficult for traditional computer vision techniques to capture the motion features of a moving object from dynamic visual scenes. Fortunately, the nature gives us many bio-visual processing inspirations, for example animals can effectively extract and perceive external motion cues by their vision systems. This helps us solve the problem of visual motion perception.

Research on neurophysiology revealed that specific types of visual neurons could respond preferentially to specific motion patterns synthesized by three basic elements, i.e., translation, expansion/contraction and rotation/depth rotation motion [3, 4]. Therein, Maunsell and Van Essen [5] reported that translational selective neurons in the dorsal part of medial superior temporal (MSTd) of the anesthetized macaque were sensitive to translational movement; Rind and Simmons [6] discovered depth motion selective neurons in the lobula complex of the locust that could positively respond to expansion vision stimuli; Saito et al. [7–9] discovered that rotational selective neurons and depth rotation sensitive (DRS) ones in the posterior parietal association cortex (area PG) of monkeys

✉ Zhuhong Zhang
zhzhang@gzu.edu.cn

Bin Hu
bhu1@gzu.edu.cn

1 Department of Computer Science, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China

2 Department of Big Data Science and Engineering, College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China

and in the medial superior temporal area (MST) of the macaque could selectively react to *depth rotation on the horizontal plane* and the fronto-parallel rotation motion for an object. The visual properties of such neurons can play an important role in engineering motion pattern detection, e.g., the motion pattern detection of a rolling wheel.

Up to now, some computational models have been developed to detect motion patterns. However, little has been done to create computational models for depth rotation motion detection. Although neurophysiologists have discovered DRS neurons in the cerebral cortex of the primate, the underlying mechanism, which a biological vision system perceives depth rotation, still remains open, let alone systematic investigation on how to design computational models for depth rotation motion detection. Therefore, it is a still open topic for researchers to discuss bio-inspired computational models used in detecting depth rotation motion in the field of engineering from the angle of computer vision. Therein, can the functional properties of the discovered DRS neurons be simulated to develop bio-inspired neurocomputational models? If yes, can such models be used to construct artificial vision systems for depth rotation object recognition and such fault detection problems as gear and propeller rotation? Therefore, the current proposal is to discuss the problem of depth rotation perception from the angle of artificial visual neural network, after which a novel computational model is designed to detect the spatio-temporal energy change of depth rotation of an object.

It is highlighted that the main contribution of the present work involves three points: (1) a bio-inspired feedforward depth rotation perception neural network (DRPNN) is originally developed to detect the pattern of depth rotation of an moving object, and thus can be applied to depth rotation object detection; (2) DRPNN can uncover some properties of depth rotation neurons, e.g., the spatio-temporal energy change property caused by depth rotation in neurophysiology; and (3) the performance characteristics of DRPNN are sufficiently examined by means of depth rotation video sequences from different scenarios.

It is worth pointing out that DRPNN differs from any existing neural networks, and in particular our previous neural network – rotational motion perception neural network (RMPNN) [10]. The main difference between DRPNN and RMPNN contains three points: (1) RMPNN suits to the detection of rotational motion on the fronto-parallel plane, whereas DRPNN is used to perceive the spatio-temporal energy change of depth rotation of an object; (2) such two neural networks originate from different biological inspirations, in other words, RMPNN is designed based on the framework of the *locust* visual system, whereas DRPNN is developed in terms of the morphological and neurophysiological characteristics of the *mammalian* visual system; and (3) RMPNN can only detect the change of translational direction of a moving object, but DRPNN involves the spatio-temporal changes of both translation and depth motion.

The rest of this paper is organized as follows. The related work on visual motion perception is reviewed in Sect. 2. Section 3 describes the proposed neural network in detail. DRPNN's computational complexity is given in Sect. 4. Section 5 displays the whole experimental analysis. Finally, Sect. 6 concludes the current work and outlines future studies.

2 Survey of related work

Depth rotation perception aims to detect specific motion patterns, in which not only objects rotate on the plane, but also the rotational axis is perpendicular to the observer's sight axis [7, 11, 12]. Up to now, many artificial visual neural networks have been proposed for different tasks in visual perception, such as target detection and tracking [13, 14], collision detection [15, 16], human identification [17, 18], visual question answering [19], intelligent surveillance [20, 21]. However, there has been no appropriate computational model for depth rotation perception in the literature. Fortunately, many achievements, reported by electro- and neuro-physiologists can give us valuable inspirations in developing bio-inspired computational models for depth rotation detection. The related work will be summarized below.

2.1 Psychophysical depth rotation perception analysis

Compared by linear movement, depth rotation movement is more difficult to be detected [22–24]. A number of psychophysical studies have been carried out to analyze different types of depth rotation perception. Although Shulman [25] claimed that the effect of attention was related to the visual process of depth rotation perception, it is not clear what factors influence the effect of visual perception. Braunstein [22] suggested that some cues affect the psychophysical perception of depth rotation. Thereafter, he designed a mental morphological model to perceive the depth rotation pattern of a rectangle. In Braunstein's model, the angle change of a rectangle between horizontal and vertical contours is taken as an indicator used in detecting depth rotation. However, his model is only assumed to be used for a rotating trapezoid or rectangle. In the study of visual factors which triggered mental depth rotation perception, Braunstein and Petersik [26, 27] reported that the factors could be separately processed. After that, Andersen and Braunstein [24, 28]

validated that mental depth rotation perception comes from the combination of directional and depth visual cues.

2.2 Geometrical model on depth rotation

By employing projective geometry approaches, Johansson et al. [29] developed a geometrical model to simulate the change of the length and direction of a straight line in depth rotation. In their simulation experiment, a depth rotating straight line is projected onto a two-dimensional plane. They suggested that the change of the projected length of the straight line be sinusoidal when it was in depth rotation. However, they only simulated an idealized depth rotation in their experiment, regardless of the influence of motion parallax on the projection of the object in the three-dimensional space. On the basis of their studies, some works are still open. For example, (1) with the aspect of motion parallax, what should be the spatio-temporal energy changes in the retina caused by the depth rotation of an object? (2) Can the perceived motion energies form a sinusoidal curve? (3) Do the energy changes, induced by depth rotation on different planes present minor differences? About these questions, we try to find their answers in this work.

2.3 Functional response of depth rotation perception

There has been a point of view that the binocular parallax is the main factor to activate DRS neurons for perceiving depth rotation [30, 31]. However, Saito et al. [7] discovered that all of the DRS neurons in the MST area of the monkey responded strongly to monocular stimulation. They also claimed that the response differences of DRS neurons under monocular and binocular vision conditions were not very distinct [9, 11, 12], while emphasizing that the motion of a single spot in depth rotation could effectively make the DRS neurons become active. These indicate that the binocular parallax is not the key factor to excite the DRS neurons, while the directional change of motion is more important than the moving object's shape. On the basis of this observation, they presented a viewpoint that the continuous change of motion direction was the only difference to distinguish rotation from linear movement. This means that specific computational models for depth rotation perception can be constructed to perceive the change of motion status of a moving object, and meanwhile two types of motion cues, i.e., depth motion and directional translation [32], need to be extracted in the early stage of visual information processing.

2.4 Computational models on directional selectivity

It has been discovered that lobula giant movement detector (LGMD) neurons in the lobula complex have the same preferential response characteristic when an object approaches the eye of a locust [33]. Rind et al. [6, 34] presented the key features of LGMD for depth motion perception, i.e., the lateral inhibition and edge expansion of an approaching object. They proposed a neural model, i.e., LGMD-based neural network for perceiving the approaching object in the 3-D space. A substantial number of experimental results suggested that the reported LGMD models work well in the perception of an object's approaching movement [35–37]. On the other hand, it has been reported that directional selective neurons widely exist in different animal species [38, 39]. Some neurophysiological achievements revealed that asymmetric lateral inhibition underlined these neurons' directional selectivity. Specifically, in animals' retina, starburst amacrine cells are connected asymmetrically to directional selective neurons, and deliver inhibition in the null directions but not in the preferred direction [40, 41]. Based on such perception mechanisms, Yue and Rind [36] proposed a directional selective neural network (DSNN) to perceive the translational direction of a moving object on the front-parallel plane. A large number of experimental results have confirmed that DSNN is robust in the perception of an object's translational motion direction [42, 43].

3 Depth rotation perception neural network

Visual motion perception depends on hierarchical information processing. Neurophysiologists have revealed that the mammalian's vision system has a layered structure and includes five types of information processing cells, respectively, presented in the five neuropil layers, i.e., photoreceptor (P), horizontal (H), bipolar (B), starburst amacrine (S), and ganglion (G) cells [44, 45]. Each of the five layers processes its input visual signals and extracts motion cues sequentially. The process of motion perception in the mammalian visual neural system can be divided into two stages [46, 47]: (1) in the first stage, motion sensitive neurons capture and transmit local motion cues to the subsequent functional neurons, and (2) in the second stage, the functional neurons with large receptive fields synthesize the received cues in order to respond to specific complex motion patterns. Inspired by such two stages of biological visual information processing, the current neural network (DRPNN) consists of a presynaptic network and a postsynaptic one. The former comprises eight lateral

inhibition neural sub-networks used for capturing visual motion information; the latter extracts the motion cues of different motion patterns, e.g., depth rotation and translational motion, and then synthesizes them to perceive the spatio-temporal energy change of depth rotation of an object.

DRPNN takes each image frame as its input signal through a monocular video camera, and then outputs the total of membrane potentials produced by its internal structures. Based on the interior characteristics of the mammalian vision system [2, 44, 45], the framework of DRPNN is developed, schematically illustrated by Fig. 1. It comprises of two parts: presynaptic and postsynaptic networks, for which the design details are given below.

3.1 Presynaptic network

In the presynaptic network of DRPNN, a depth perception (*DP*) neuron is used to capture the approaching/receding cues of depth motion of an object, while eight directional selective neurons are utilized to extract various translational direction cues of the object. On the basis of the preferred translational motion directions, the eight directional selective neurons, which correspond to respective directional selective neural networks (DSNNs), can be

Fig. 1 Schematic illustration on DRPNN

classified into two types. The first type, which consists of horizontal and vertical directional selective neurons, includes left (L), right (R), up (U), and down (D) selective neurons; the second type, which is formed of diagonally directional selective neurons, involves in left-up (LU), left-down (LD), right-up (RU), and right-down (RD) selective neurons. Each of the eight neurons corresponds to a special DSNN which perceives specific translational direction cues of the object. Therefore, the presynaptic network includes eight DSNNs acquired by improving the reported computational models [36, 37]. Such eight neural networks share the four layers of *P*, *H*, *B* and *S*, but have different designs for their *G* layers, i.e., their direction inhibition layers. We take the left directional selective neural network(L-DSNN) for example to illustrate their internal frameworks and functional mechanisms.

As shown in the top part of Fig. 1, L-DSNN, which preferentially responds to a left moving object in the field of view, includes five neural information-processing layers and one functional neuron, i.e., *P*, *H*, *B*, *S* and *G* layers, and the mentioned left selective neuron (*L*). The functions of each neural layer and neuron *L* are described below.

1) *P* layer

The *P* layer as the first layer of L-DSNN is to capture the visual motion signals of an object in the field of view. By an analogy to the morphologic characterization of the mammalian's retina [44, 45, 48], it consists of $n_c \times n_r$ photoreceptor cells arranged in a matrix form. Each cell receives the luminance intensity or gray value at the counterpart in an input image frame with size $n_c \times n_r$. Let *x* and *y* denote the row and column coordinates in the input image, respectively. L_{f-1} and L_f are the luminance values at frames *f* and *f* − 1, respectively. $P_f(x, y)$ represents the captured luminance change which corresponds to pixel (*x*, *y*) at frame *f*, given [35] by

$$P_f(x, y) = \text{abs}(L_f(x, y) - L_{f-1}(x, y)). \tag{1}$$

By the neurophysiologic achievements of the mammalian vision system [49, 50], the output of cell (*x*, *y*), $\hat{P}_f(x, y)$, is determined by

$$\hat{P}_f(x, y) = \begin{cases} P_f(x, y), & \text{if } P_f(x, y) \geq T_{rp}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

with signal threshold T_{rp}.

Remark 1 To clarify the functionality of the *P* layer, a video sequence is utilized to show a left moving ball on a carpeted office room (see Fig. 2a). We take two successive image frames with numbers 48–49 for example to illustrate the processed result. After receiving frame 49, the *P* layer firstly computes the changes of luminance intensities in terms of frame 48 and Eq. (1), by which the changes form a difference image given in Fig. 2b. We notice that the
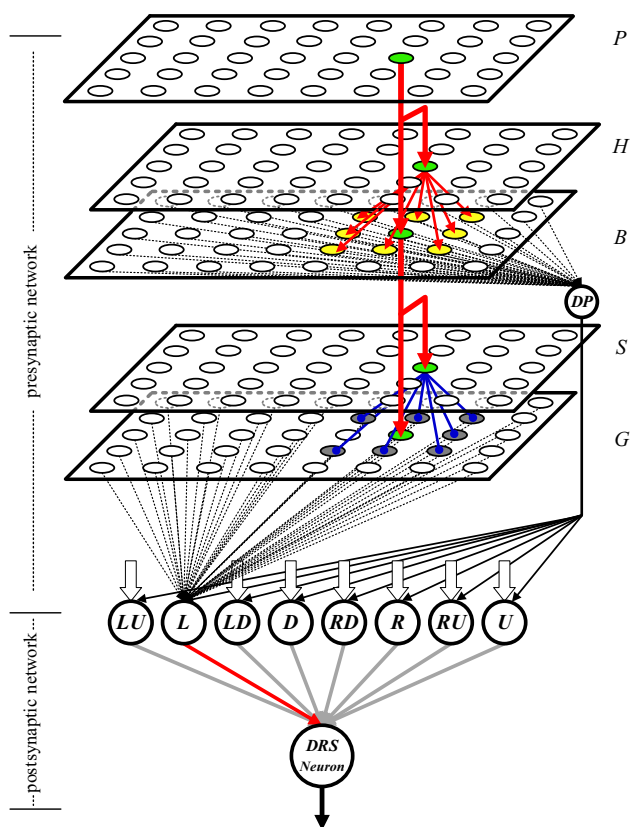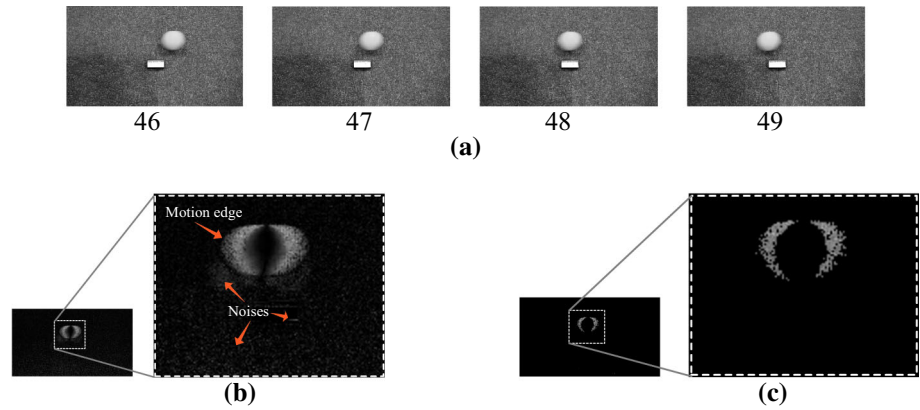
Fig. 2 Illustrative example on the P layer

motion edge of the moving ball can be extracted, but some noises are included. Subsequently, the image is transformed into Fig. 2c after being further processed by means of Eq. (2) and a given threshold value of T_{rp}.

2) H and B layers

The H and B layers as the second and the third layers of L-DSNN, respectively, are designed based on the mammalian's neurophysiologic findings, namely that the horizontal cells can collect visual signals from the photoreceptor cells and provide feedforward signals to the bipolar cells for the improvement of the spatial resolution of visual information [48, 51, 52]. Each of such two layers includes $n_c \times n_r$ cells displayed in a matrix form. Those cells in the H layer directly receive the excitatory intensities from their retinotopic counterparts in the P layer, and then transmit them to the B layer. Therein, the output of each cell (x, y) is defined by

$$H_f(x, y) = \hat{P}_f(x, y). \tag{3}$$

In the B layer, each cell not only collects the outputs of the cells around its retinotopic counterpart in the H layer, but also fuses such outputs with the output from its retinotopic counterpart in the P layer. More precisely, as related to the visual information integration metaphor in the mammalian's retina [51, 52], the surrounded excitation of each cell from the H layer to the B layer can only get a smaller passing coefficient of w_{hb}, and conversely the direct excitation of each cell from the P layer to the B layer gains a larger passing coefficient of w_{pb} in the process of information integration. Therefore, the strength of the mixed excitation $B_f(x, y)$ of each cell in the B layer is given through

$$B_f(x, y) = \sum_{i=-m_w}^{m_w} \sum_{j=-m_w}^{m_w} H_f(x+i, y+j)w_c(i, j))w_{hb} + \hat{P}_f(x, y)w_{pb}, \tag{4}$$

with surround radius m_w, where w_c denotes a convolution mask given by

$$w_c = \begin{bmatrix} 0.125 & 0.25 & 0.125 \\ 0.25 & 0 & 0.25 \\ 0.125 & 0.25 & 0.125 \end{bmatrix}, \tag{5}$$

based on the neurophysiological achievements [52, 53] and empirical experience [20, 36, 54].

Remark 2 As related to Fig. 2c, the B layer generates $n_c \times n_r$ visual excitations to form an image given in Fig. 3, relying upon the H layer and Eq. (4). Figure indicates that the object's motion edge becomes clearer and some clutters in Fig. 2c can be filtered out.

3) S and G layers

Neurophysiological studies have revealed that starburst amacrine cells as inhibitory inter-neurons play an important role in forming the visual perception of directional selectivity. More precisely, such cells gather signals from the bipolar cells and passes their directional inhibition signals to the ganglion cells in the null directions but not in the preferred direction by means of their major synapses [2, 44]. Analogously, the fourth and fifth layers of L-DSNN are the S and G layers, respectively, each of which is arranged in a $n_c \times n_r$ matrix form. Each cell in the S layer receives the membrane potential of its retinotopic counterpart in the B layer, and generates its left inhibition through
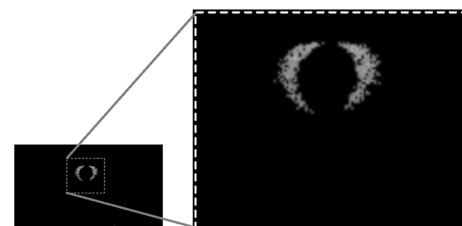

Fig. 3 Illustrative image frame acquired by the B layer

$$I_f^{\mathrm{L}}(x,y) = \sum_{i=-n_{\mathrm{inh}}}^{n_{\mathrm{inh}}} \sum_{j=-n_{\mathrm{inh}}}^{n_{\mathrm{inh}}} B_{f-1}(x+i, y+j)$$
$$- \sum_{i=1}^{n_{\mathrm{inh}}} B_{f-1}(x+i,y) w_\xi(i), \qquad (6)$$

where the superscript L denotes the L neuron; n_{inh} is the inhibition radius; $w_\xi(i)$ is the local inhibition weight which controls the opposite side neighboring inhibition strength given by

$$w_\xi(i) = (2n_{inh} + 1)^2 \psi_I, \qquad (7)$$

with membrane potential constant ψ_I. Then, each cell in the S layer outputs its inhibition intensity [35] through

$$\hat{I}_f^{\mathrm{L}}(x,y) = \begin{cases} I_f^{\mathrm{L}}(x,y), & \text{if } I_f^{\mathrm{L}}(x,y) > 0, \\ 0, & \text{else.} \end{cases} \qquad (8)$$

Additionally, as related to the mammalian's neurophysiological achievement which the intensities of excitation and inhibition from bipolar and starburst amacrine cells need to pass to ganglion cells [2, 44], each cell in the G layer collects two types of visual signals of the above B and S layers. One is the output excitation from the retinotopic counterpart in the B layer, and the other is the gathered inhibition spread by the retinotopic counterpart's neighboring cells in the S layer. The collected visual signals are integrated by

$$G_f^{\mathrm{L}}(x,y) = B_f(x,y) - \hat{I}_f^{\mathrm{L}}(x,y) w_I, \qquad (9)$$

where $G_f^{\mathrm{L}}(x,y)$ is the integrated excitation of cell (x, y) in the G layer; w_I is the global inhibition weight which controls the whole inhibition strength. Subsequently, only those cells whose membrane potentials exceed a threshold value of T_g will output their activities. Therefore, if the membrane potential of a cell in the G layer is smaller than T_g, its output is set as 0, and remains unchanged otherwise. The output of each cell in the G layer is computed by

$$\hat{G}_f^{\mathrm{L}}(x,y) = \begin{cases} G_f^{\mathrm{L}}(x,y), & \text{if } G_f^{\mathrm{L}}(x,y) \geq T_g, \\ 0, & \text{else.} \end{cases} \qquad (10)$$

Remark 3 The S and G layers in L-DSNN are utilized to extract the directional visual cues of a moving object. Figure 4a, b presents the outputted inhibitions and excitations of cells in the S and G layers, i.e., those excitations at
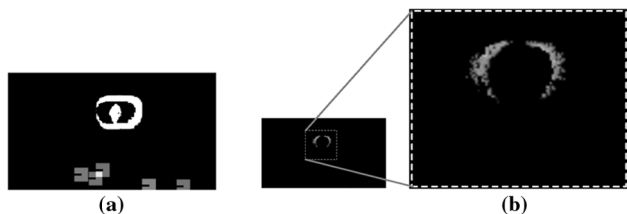

Fig. 4 The outputted frames: **a** the S layer; **b** the G layer

frame 49, respectively. The directional inhibitions from the S layer are allocated to the cells in the G layer in the null directions but not in its preferred direction.

4) Neuron L

The output membrane potentials of all cells in the G layer are gathered to the L neuron. The strength of the converged excitation is computed [35] by

$$\mathrm{SUM}_f^{\mathrm{L}} = \sum_{x=1}^{n_c} \sum_{y=1}^{n_r} \mathrm{abs}(\hat{G}_f^{\mathrm{L}}(x,y)). \qquad (11)$$

Then, the acquired excitation is given by

$$E_f^{\mathrm{L}} = 2 \times \left(1 + e^{-\frac{\mathrm{SUM}_f^{\mathrm{L}}}{n_r n_c}}\right)^{-1} - 1, \qquad (12)$$

where E_f^{L} is the output excitation of the neuron L, and it changes within 0 and 1, due to $\mathrm{SUM}_f^{\mathrm{L}} \geq 0$.

5) Other directional selective neural networks

Besides the above L-DSNN, DRPNN includes seven subnetworks related to the corresponding neurons, e.g., R-DSNN, LU-DSNN, etc. These DSNNs have the same designs as those in the L-DSNN except their directional inhibition designs in their G layers. Here, we only take LU-DSNN for example to illustrate their inhibition gathering designs. The gathered inhibition strength of each cell (x, y) in the G layer is defined by

$$I_f^{\mathrm{LU}}(x,y) = \sum_{i=-n_{\mathrm{inh}}}^{n_{\mathrm{inh}}} \sum_{j=-n_{\mathrm{inh}}}^{n_{\mathrm{inh}}} B_{f-1}(x+i, y+j)$$
$$- \sum_{j=1, i=j}^{n_{\mathrm{inh}}} B_{f-1}(x+i, y+j) w_\xi(i,j). \qquad (13)$$

6) Depth perception neuron—DP neuron

The DP neuron corresponds to the depth perception neural network used for capturing the approaching/receding cues of a moving object in depth motion. The DP neuron and the eight directional selective neurons share the same neural layers (i.e., P, H and B layers) of the presynaptic network in processing visual signals, as shown in Fig. 1 (top). It gathers the output excitations of all cells in the B layer, and gathers their excitations by [35]

$$\mathrm{SUM}_f^{\mathrm{DP}} = \sum_{x=1}^{n_c} \sum_{y=1}^{n_r} \mathrm{abs}(B_f(x,y)). \qquad (14)$$

After that, the DP neuron's output is decided by

$$E_f^{\mathrm{DP}} = 2 \times \left(1 + e^{-\frac{\mathrm{SUM}_f^{\mathrm{DP}}}{n_r n_c}}\right)^{-1} - 1, \qquad (15)$$

where E_f^{DP} is the output excitation of the DP neuron at frame f.

Remark 4 In order to demonstrate the functionality of the *DP* neuron in capturing depth motion cues, a video sequence is chosen to illustrate how a ball approaches the video camera. Four of all the video frames, presented by Fig. 5a are picked up to represent such an approaching motion pattern. After the video frames are orderly processed by Eqs. (1)–(5), (14) and (15), the *DP* neuron generates an output curve given in Fig. 5b. We note that the output excitation of the *DP* neuron will become large when the ball approaches the camera increasingly. This indicates that the *DP* neuron can correctly perceive the depth motion cues caused by the approaching ball, and thus possesses the perception feature of depth motion [34, 55].

3.2 Postsynaptic network

The postsynaptic network receives the extracted visual motion cues from the above presynaptic network for further processing. As shown in Fig. 1 (bottom), it consists of a neuropil layer and a functional neuron, i.e., the direction column and *DRS* neuron.

　1)　Direction column

Inspired by a neurophysiological finding that the mammalian cerebral cortex includes visual neurons with respective motion preference axes and exists in the form of direction column [56, 57], the above-mentioned eight directional selective neurons form a direction column. According to the ranked order of the neurons, the excitations of the neurons are represented by the following expression,

$$\Psi_f = (E_f^{\mathrm{LU}}, E_f^{\mathrm{L}}, E_f^{\mathrm{LD}}, E_f^{\mathrm{D}}, E_f^{\mathrm{RD}}, E_f^{\mathrm{R}}, E_f^{\mathrm{RU}}, E_f^{\mathrm{U}}). \quad (16)$$

A spiking mechanism is utilized to determine the values of elements in ψ_f. More precisely, an internal spike $\kappa_f(i)$ occurs inside this neuron i with $i \in \{LU,L,\ldots,RU,R\}$, that is

$$\kappa_f(i) = \begin{cases} 1, & \text{if } \Psi_f(i) \geq T_e \wedge \Psi_f(i) > 0, \\ 0, & \text{else}, \end{cases} \quad (17)$$

where $T_e = \max\{\Psi_f(i), \ 1 \leq i \leq 8\}$. If n_s successive spikes occur, the output excitation of the neuron i is computed by

$$\hat{\Psi}_f(i) = \begin{cases} (-1)^{(\mathrm{Quotient}(i,\lambda_d)+1)} \times \mathrm{E}_f^{\mathrm{DP}}, & \text{if } \sum_{m=f_s^i}^{f} \kappa_m(i) \geq n_s, \\ 0, & \text{else}, \end{cases} \quad (18)$$

with quotient function *Quotient*(.,.) and constant divisor λ_d, where f_s^i denotes the first frame of the time period when continuous spikes are occurring inside the directional selective neuron i; the threshold n_s is defined by

$$n_s = \max\left\{\sum_{m=f_s^i}^{f} \kappa_m(i), \quad 0 \leq i \leq 8\right\}. \quad (19)$$

As the processing mechanism was described above, all elements in the output excitation vector $\hat{\Psi}_f$ are zero if the object keeps static, and conversely at least one element will be larger than zero.

Remark 4 With the unique network structure of the direction column, the current motion direction of a moving object can always be detected by the eight neurons. Figure 6 exhibits the excitation curves of the neurons in terms of the video sequence in Remark 1. We notice that the *LU*, *L* and *LD* neurons become excitatory, since their excitation curves beyond those of the other neurons. This is in accordance with the fact that the object moves on the left direction.

　2)　Depth rotation sensitive neuron—*DRS* neuron

The outputs of the above eight neurons in the direction column are converged to the DRS neuron, by which the strength of the membrane potential of it at frame f is computed by

$$E_f^{\mathrm{DRS}} = \sum_{i=1}^{8} \hat{\Psi}_f(i), \quad (20)$$

After that, one such neuron produces its membrane potential taken as the output of DRPNN.

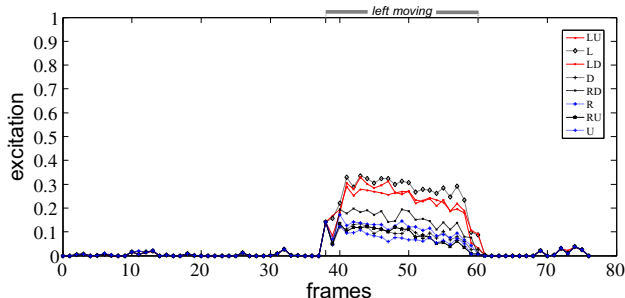**Fig. 5** Illustrative example on depth motion perception

(a)

(b)
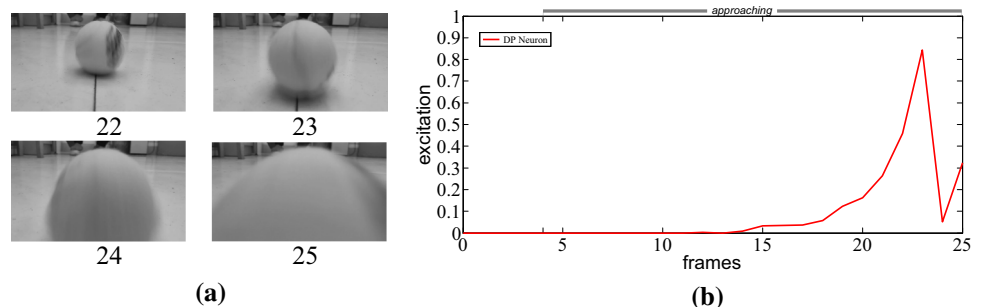
Fig. 6 Excitation curves acquired by the neurons

4 Computational complexity

Let N be the total of pixels of each inputted image frame with $N = n_c \times n_r$. Within an iterative cycle, DRPNN executes $3N$ operations in the P layer, while the H and B layers involve in 21 arithmetic operations. The S layer only enforces N assignment operations. The G layer needs M_1 times to extract motion cues in the eight translational directions with $M_1 = (32n_{inh}^2 + 40n_{inh} + 72)N$. Additionally, the eight directional selective neurons are required to run $16(N + 3)$ arithmetic operations, while the DP neuron executes $2N + 7$ operations. Furthermore, in the direction column there needs to operate M_2 operations with $M_2 = 8(f - f_s^i) + 68$. Finally, the DRS neuron enforces 7 addition/ subtraction operations. Summarily, the total of DRPNN's executions for a loop is decided by

$$\text{Sum} = (32n_{inh}^2 + 40n_{inh} + 94)N + 8(f - f_s^i) + 151. \quad (21)$$

Since $f - f_s$ takes small values, DRPNN's computational complexity in the worst case is given by

$$O = O(32n_{inh}^2 + 40n_{inh} + 94)N. \quad (22)$$

Equation (22) shows that the image resolution N and the inhibition radius n_{inh} influence DRPNN's computational efficiency. Therefore, it will be beneficial to reduce the input video frame size and take a rational inhibition radius value in the G layer.

5 Experimental study

In the study of the DRS neurons in macaque's cerebral cortex, it was found that these visual neurons can perceive the depth rotation in the field of view [7, 9, 11, 12]. Therefore, we use several sets of video sequences to analyze the performance of DRPNN. More specifically, in order to check whether DRPNN can effectively and robustly perceive the spatio-temporal energy change of depth rotation and also whether its output curve is a sinusoidal curve or not, several real scenarios, which reflect the

specific depth rotation of a moving object, are firstly set to sample video sequences; secondly, DRPNN is sufficiently verified, which involves depth rotation on the horizontal and non-horizontal planes; finally, it is compared by three recent motion perception neural networks. The architecture of the experiment flowchart is given in Fig. 7.

5.1 Experimental environment

All experiments are executed on a Microsoft Windows 10 computer with CPU/2.66G and RAM/4G by means of VC++ platform. Thirty-four video sequences are taken to examine the performance of DRPNN. Each video sequence is recorded at a frame rate of 30fps, and later separated into 8-bit grayscale images with size 140×80 per frame.

The parameter settings of DRPNN are given in Table 1. n_c and n_r are set as 140 and 80, respectively, since cells in the P layer correspond to the pixels in the input image. The potential constant ψ_I is set as 255, based on the maximum value of the pixel in the 8-bit grayscale image. The proportion weights w_{hb} and w_{pb} are defined as 0.33 and 0.67, respectively, which bases on the visual information integration metaphor in the mammalian's retina [51, 52]. m_w, w_I, n_{inh}, and T_g take 1, 1.7, 4 and 12, respectively, which depends on the previous experiments [10, 20, 36, 37, 54].

5.2 Depth rotation perception on the horizontal plane

In the study of DRS neurons in MSTd, it was found that such neurons can well respond to depth rotation on the horizontal plane [7, 9, 11, 12]. Therefore, from the angle of computer simulation we test whether DRPNN can simulate one such property by means of a set of video sequences which represent the horizontal depth rotation of a rigid object. More precisely, when the object is in depth rotation, there are two types of projection shapes in the retina, i.e., non-deformation and deformation. Hereby, two regular ball and rectangle are used to generate four video sequences used for detecting whether DRPNN can well respond to objects' depth rotation. The schematics of four typical depth rotation patterns on the horizontal plane are shown in Fig. 8.

In terms of a monocular video camera, we firstly record two video sequences produced by a regular black ball (40 mm in diameter) that rotates around a fixed rotation center on the horizontal plane. Similarly, we also record two video sequences generated by a regular black rectangle (80 mm in length and 35 mm in width) that rotates around one of its fixed edges on the horizontal plane. In each video sequence, the object is placed in the central region of the field of view and rotates at a constant angular velocity (see Fig. 9). Depending on the four groups of video sequences,

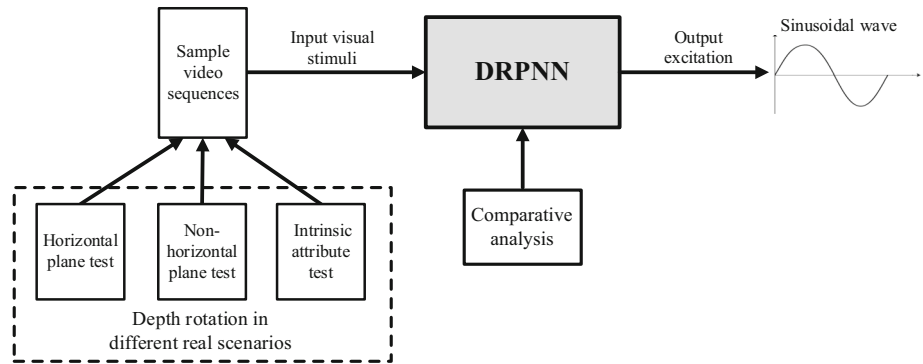**Fig. 7** Schematic illustration on the architecture of the experimental study

Table 1 Parameter settings of DRPNN

Name	Value	Name	Value
Signal threshold T_{rp}	50	Potential constant ψ_I	255
Surround radius m_w	1	Inhibition weight w_I	1.7
Proportion weight w_{hb}	0.33	Signal threshold T_g	12
Proportion weight w_{pb}	0.67	Resolution $n_c \times n_r$	140×80
Inhibition radius n_{inh}	4	Constant divisor λ_d	4

we verify whether DRPNN can effectively perceive the motion change of an object in depth rotation and whether its output excitation presents a sinusoidal curve or not.

In Fig. 9a, a black ball is at the leftmost position of the rotation trajectory and keeps stationary from frame 1 to frame 31, and later it rotates counterclockwise one circle at an angular velocity (6.5 rad/s) on the horizontal plane from frame 32 to frame 60; finally, it remains stationary from frame 61 to frame 88. In Fig. 9b, the black ball is at the rightmost position of the rotation trajectory and then holds stationary from frame 1 to frame 33; subsequently, it rotates clockwise one circle at an angular velocity (6.5 rad/s) on the horizontal plane from frame 34 to frame 61; finally, it holds stationary from frame 62 to the end. Similarly, in Fig. 9c, d, the depth rotation pattern of a rectangle is similar to that of the ball shown in Fig. 9a or b except that the angular velocity is 8.56 rad/s. The statistical results, acquired by DRPNN are displayed in Table 2, and meanwhile Fig. 10 presents DRPNN's output curves as related to Fig. 9.

By Fig. 10 and Table 2, when an object is in depth rotation, DRPNN can be touched to respond to the spatio-temporal energy changes occurring in the field of view and outputs its excitation. However, its output curve is somewhat different from a standard sinusoidal one in this test. In other words, (1) the left and right sub-parts of it are asymmetry; (2) the absolute values of the up and down

peaks of the curve are not equal; and (3) there are some perturbations presented in the peaks. The main reason is because depth rotation easily causes motion parallax. Herein, four conclusions can be drawn: (1) DRPNN can effectively perceive the depth rotation of a moving object on the horizontal plane, regardless of the object's shape; (2) the spatio-temporal energy change, caused by depth rotation can be effectively captured by DRPNN; and (3) the membrane potential curve, outputted by DRPNN presents a quasi-sinusoidal curve which is compatible with the hypothesis suggested by Johansson et al. [29] in projective geometry.

5.3 Depth rotation perception on the non-horizontal plane

This section detects whether DRPNN can perceive the change of spatio-temporal energy if depth rotation takes place on the non-horizontal plane [9, 11, 12]. Herein, a monocular video camera records six video sequences which arise from the depth rotation of a ball on the different non-horizontal planes. Three typical non-horizontal planes, i.e., the left diagonal, the sagittal, and the right diagonal planes, are employed to produce video sequences. The schematics of depth rotation on the three typical non-horizontal planes are shown in Fig. 11 above.

As related to Fig. 11, Fig. 12 presents six video sequences that depict different kinds of depth rotation patterns of a ball. In Fig. 12, the depth rotation patterns of the ball are similar to that of the above depth rotation on the horizontal plane test. The experimental results can be known by Fig. 13.

Figure 13 indicates that in the case of non-horizontal depth rotation, DRPNN can also perceive the spatio-temporal energy change of depth rotation of the object. Particularly, the output curves of DRPNN are also quasi-sinusoidal, and thus DRPNN can simulate the property of which the DRP neurons can perceive the depth rotation of an object on a non-horizontal plane [9, 11, 12].

Fig. 8 Schematic examples of four typical depth rotation patterns on the horizontal plane. The object rotates on the horizontal plane (i.e., the *X–Z* plane). The sight axis of the video camera coincides with the *Z* axis and is perpendicular to the rotation axis of the object (i.e., the *Y* axis). The red direction line indicates the rotation direction of the object: **a** the ball is in counterclockwise (ccw) depth rotation, **b** the ball is in clockwise (cw) depth rotation, **c** the rectangle is in ccw depth rotation, and **d** the rectangle is in cw depth rotation
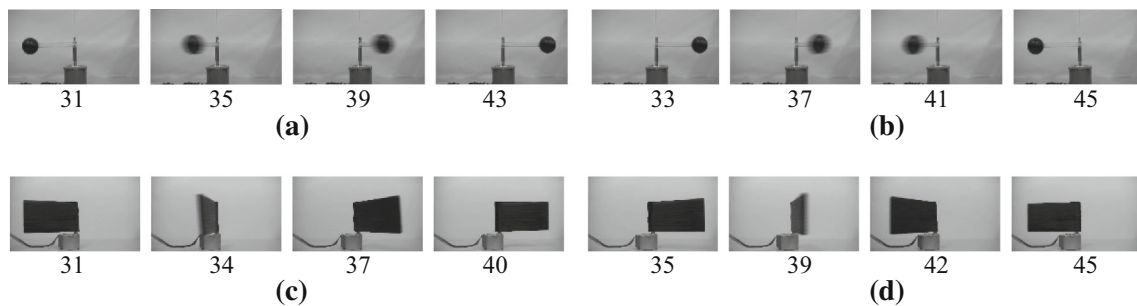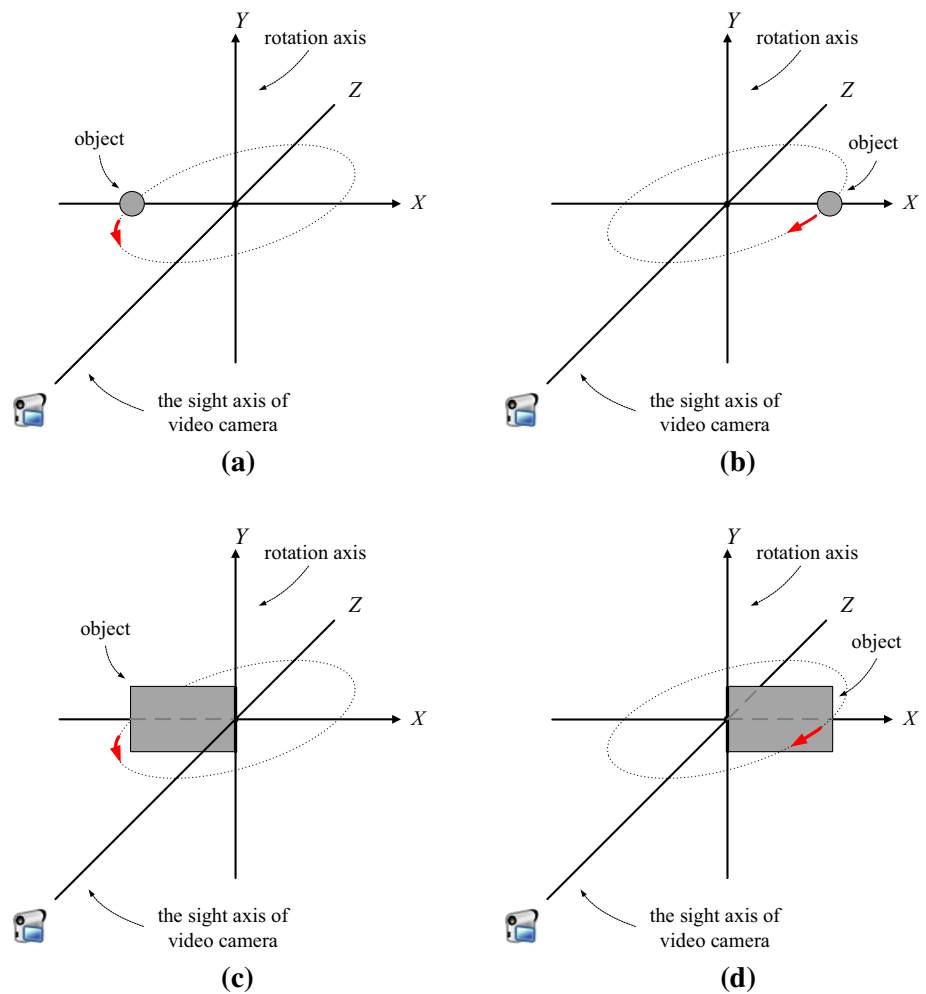
Fig. 9 As related to Fig. 8, the example frames of depth rotation on the horizontal plane are given here. Each video sequence is illustrated only by picking up four frames; the frame number is indicated under each image. **a** ccw depth rotating ball, **b** cw depth rotating ball, **c** ccw depth rotating rectangle, and **d** cw depth rotating rectangle

5.4 DRPNN's intrinsic property

5.4.1 Case I: Position invariance test

Four video sequences in Fig. 14 are sampled based on the scenarios of horizontal depth rotation for a ball. Each of them is gotten on a specific non-central region of the field of view (i.e., top-left, bottom-left, top-right, and bottom-

right); the ball with angular velocity 6.49 rad/s makes depth rotation, while its depth rotation pattern is similar to that of the ball shown in Fig. 9a. As related to the video sequences in Fig. 14, DRPNN is executed on each video sequence. Its output curves are displayed in Fig. 15.

The curves show that DRPNN can correctly perceive the motion change of depth rotation while outputting quasi-sinusoidal curves, even if the current depth rotation occurs

Table 2 Depth rotation perception region and experimental results gotten by DRPNN in the horizontal plane test

Video	The total number of frames	Object type	Rotating plane of the object	Rotation direction	Practical depth rotation region (frames)	Angular velocity (rad/s)	DRPNN's perception region (frames)	Perception success rate (%)	Perception false rate (%)
a	88	Ball	Horizontal	*ccw*	32–60	6.5	32–60	100	0
b	91	Ball	Horizontal	*cw*	34–61	6.5	34–61	100	0
c	79	Rectangle	Horizontal	*ccw*	31–52	8.56	31–52	100	0
d	83	Rectangle	Horizontal	*cw*	36–57	8.56	36–57	100	0
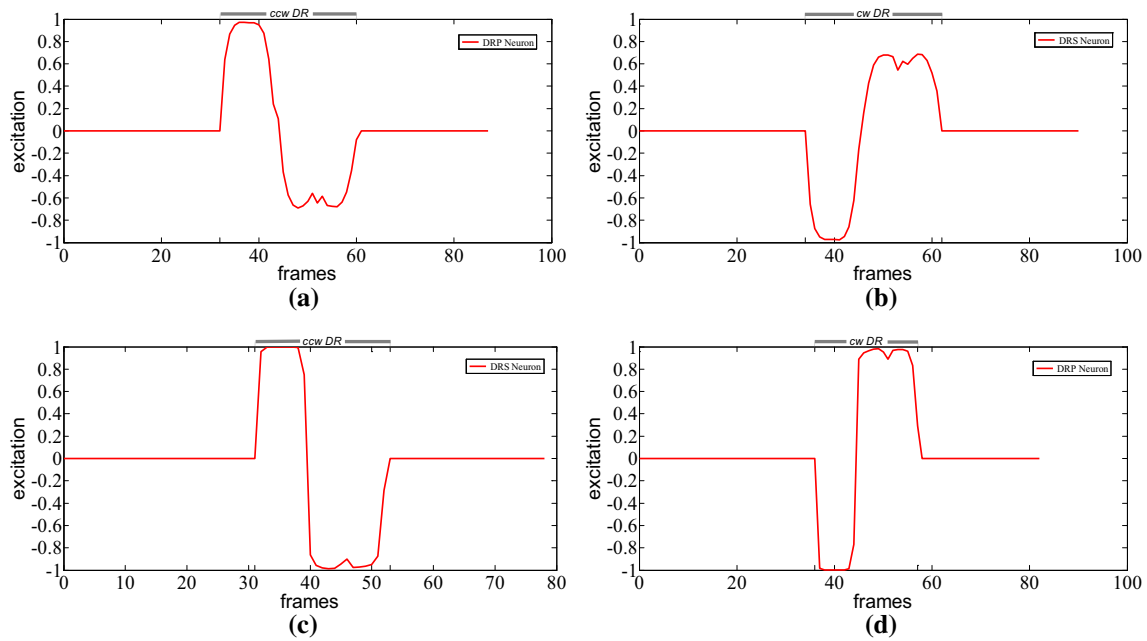
Fig. 10 Output curves of DRPNN. The four subfigures are acquired by the above corresponding video sequences in Fig. 9
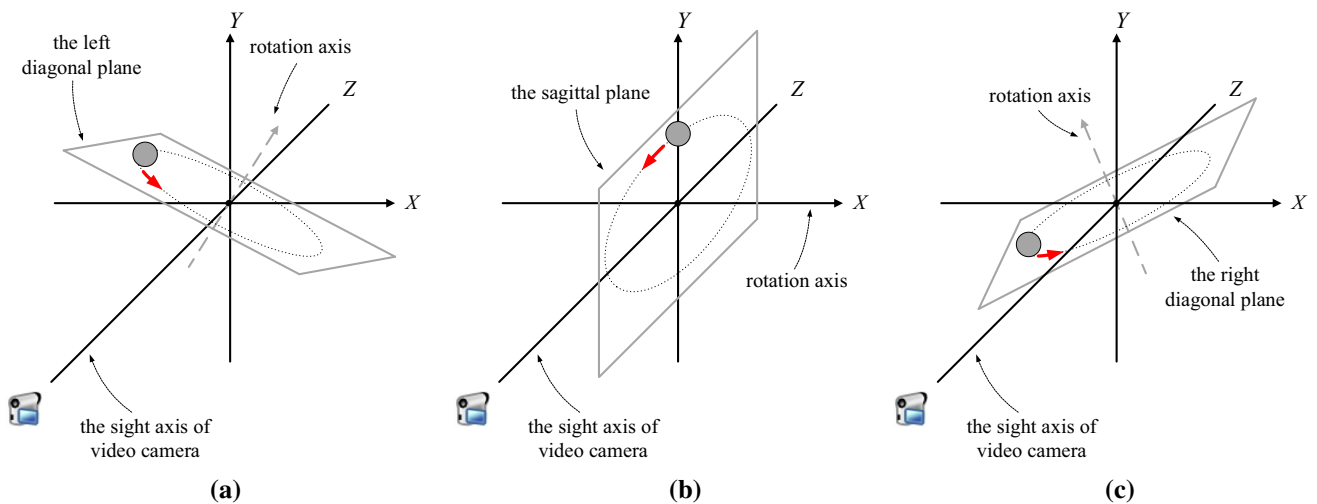
Fig. 11 Schematic illustrations of depth rotation on the three typical non-horizontal planes. **a** ccw depth rotation on the left diagonal plane; **b** ccw depth rotation on the sagittal plane; **c** ccw depth rotation on the right diagonal plane
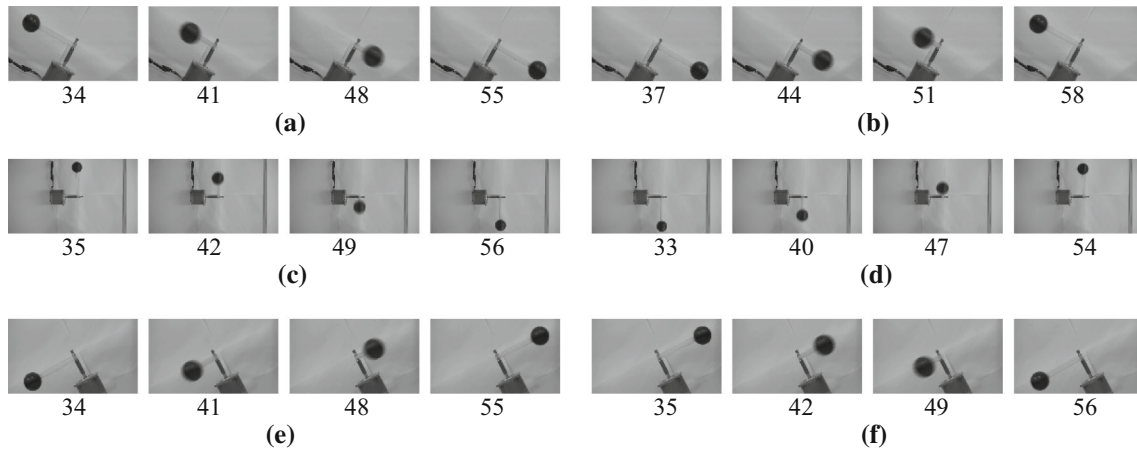
Fig. 12 Example frames based on the different rotating planes. Each video sequence is illustrated only by picking up four frames; the frame number is indicated under each image. **a** ccw depth rotation on the left diagonal plane; **b** cw depth rotation on the left diagonal plane; **c** ccw depth rotation on the sagittal plane; **d** cw depth rotation on the sagittal plane; **e** ccw depth rotation on the right diagonal plane; **f** cw depth rotation on the right diagonal plane
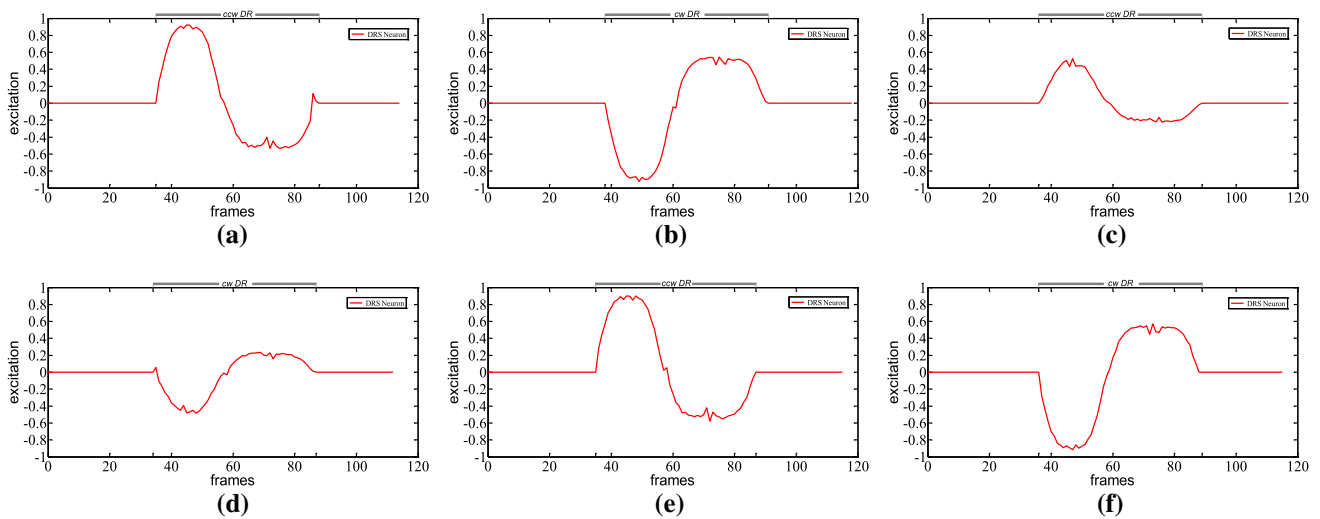
Fig. 13 Output curves of DRPNN. Each subgraph uniquely corresponds to the experimental result of the same identifier video sequence in Fig. 12
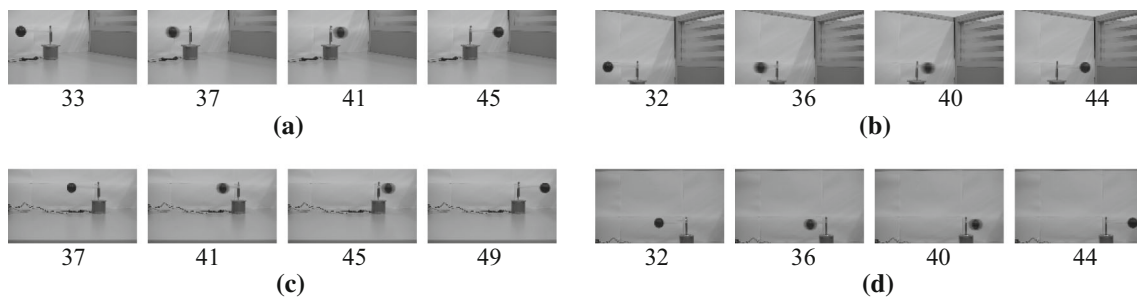
Fig. 14 Sample frames with different scenarios. Each video sequence is represented with only four frames; the frame number is indicated under each image. **a** top-left region, **b** bottom-left region, **c** top-right region, and **d** bottom-right region
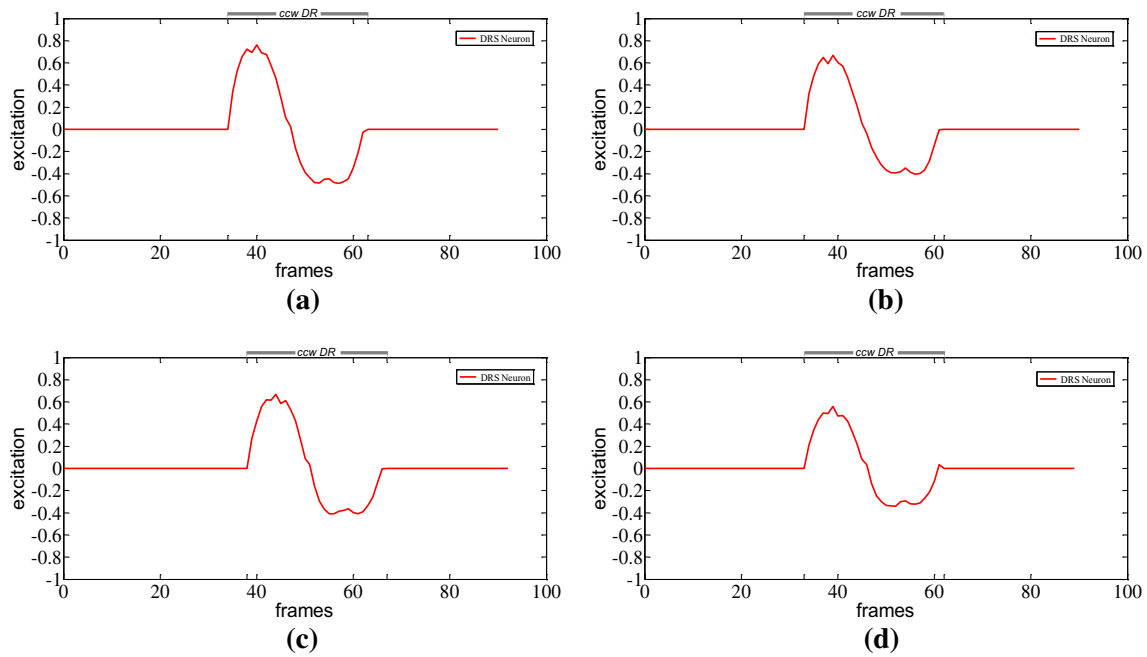
Fig. 15 Output curves of DRPNN. Each subgraph is acquired by DRPNN through the corresponding video sequence in Fig. 14

within different non-central regions of the field of view. This is consistent with the property of DRP neurons, namely they can make excitation wherever depth rotation takes place [7, 11].

5.4.2 Case II: Sensitivity on rotation speed

We here examine how the output curve of DRPNN is influenced by different rotation speeds. Here, six video sequences, generated from the horizontal depth rotation of a ball with different angular velocities are taken (see Fig. 16). The depth rotation patterns in Fig. 16, orderly

with rotation angular velocities 1.81, 2.27, 3.55, 9.92, 20.93 and 47.1 rad/s, are similar to that in Fig. 9a.

Figure 17 displays the output excitation curves of DRPNN, which hints that DRPNN can perceive the depth rotation of the moving object. We also observe that, when the rotation angular velocity of the rotating object is smaller than or equal to 9.92 rad/s (see Fig. 17a–d), the output curve of DRPNN is a quasi-sinusoidal curve. However, when the rotation speed is equal to or larger than 20.93 rad/s, the output curve of DRPNN is not similar to a quasi-sinusoidal curve. This shows that an appropriate rotation speed can help DRPNN correctly perceive the pattern of depth rotation.
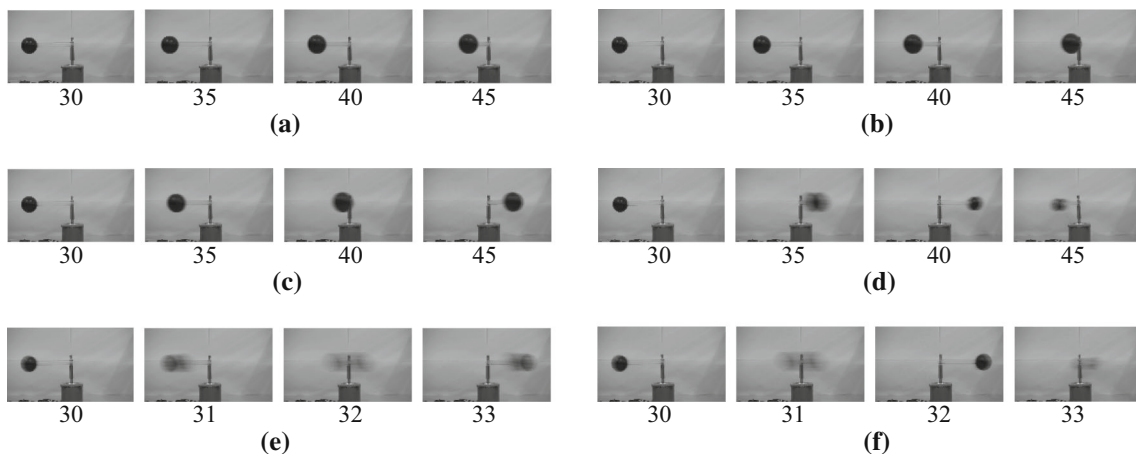

Fig. 16 Example frames with different rotation angular velocities. Each video sequence is illustrated only by picking up four frames; the frame number is indicated under each image
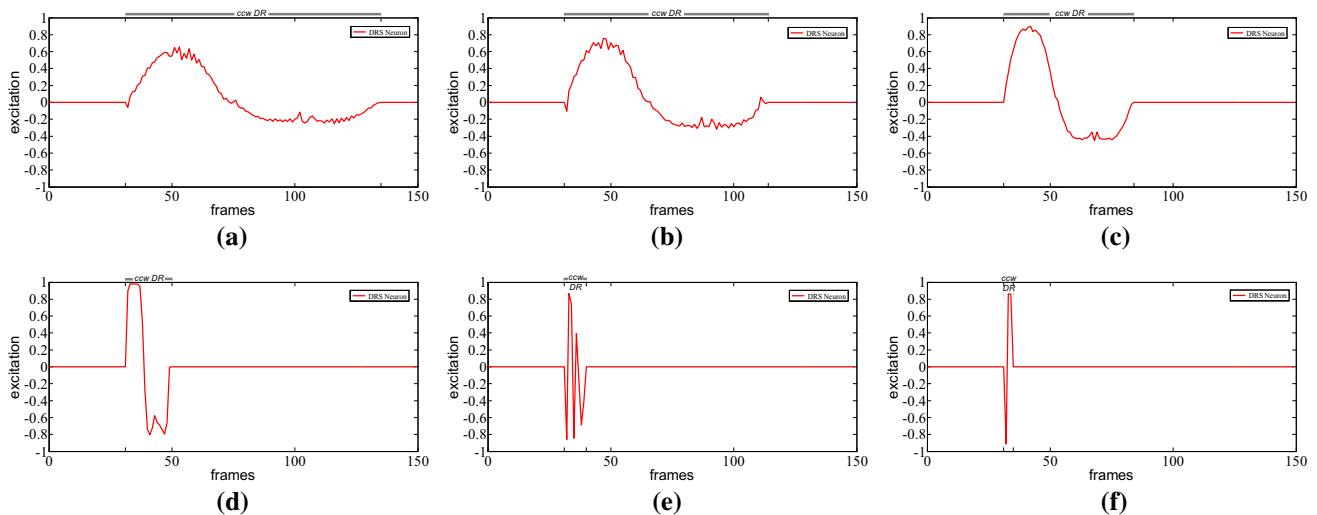
Fig. 17 Output curves of DRPNN. Each subgraph is acquired by DRPNN through the corresponding video sequence in Fig. 16

5.4.3 Case III: Sensitivity on the starting point

In the case II, the starting point of the rotating object is at the leftmost or rightmost position of the rotation trajectory, and thus DRPNN can produce a quasi-sinusoidal excitation curve when the object is in depth rotation. In order to confirm that depth rotation may happen through other starting points in its rotation trajectory, we use a set of video sequences, arisen from the horizontal depth rotation of a ball with different starting points to challenge DRPNN. In these video sequences, the ball rotates counterclockwise on the horizontal plane (i.e., the X–Z plane) with different and specific starting points. The schematic illustration of the test scenarios is given in Fig. 18.

As illustrated by Fig. 18, the first starting point (P_1) is at the leftmost position of the rotation trajectory of the ball. Then, the horizontal deviation angle between the next starting point (P_2) and P_1 increases $45°$. Similarly, the
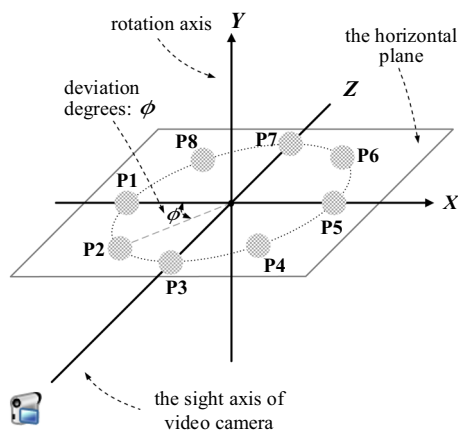

Fig. 18 Schematic illustration based on different starting points

angle between the kth starting point (P_k) and P_1 is $k \times 45°$ with $1 < k \leq 8$. Hence, the eight test video sequences acquired in Fig. 19 are used to formulate the process of depth rotation of the ball in terms of the eight starting points. In the video sequence of Fig. 19a with a total of 112 frames, the ball is located at the first starting point P_1 and keeps stationary from frame 1 to frame 30, after which it rotates counterclockwise one circle on the horizontal plane at an angular velocity 3.55 rad/s from frame 31 to frame 83, and finally it remains stationary from frame 84 to the end. In the video sequences displayed in Fig. 19b–h, the motion patterns of the ball are similar to that presented in Fig. 19a.

As related to the video sequences shown in Fig. 19, Fig. 20 displays the output curves of DRPNN. By Fig. 20a, DRPNN can perceive the motion change of depth rotation of the ball, and its output curve is a quasi-sinusoidal curve. However, with the change of the starting point of depth rotation, the phase shifts of its output waveform will take place (see Fig. 20b–h). These test results indicate that the starting point of depth rotation can influence the phase of the output waveform perceived by DRPNN. The output curve, however, is still quasi-sinusoidal.

5.4.4 Case IV: Sight axis deviation test

In the above test, the sight axis of the video camera overlaps the rotating plane of the object, in which the camera is perpendicular to the rotation axis. Herein, we examine how the sight axis deviation influences the output waveform of DRPNN. More precisely, we test how DRPNN responds to the depth rotation of an object in the case where the sight axis of the video camera deviates from the rotating plane of the object. Here, take horizontal ccw
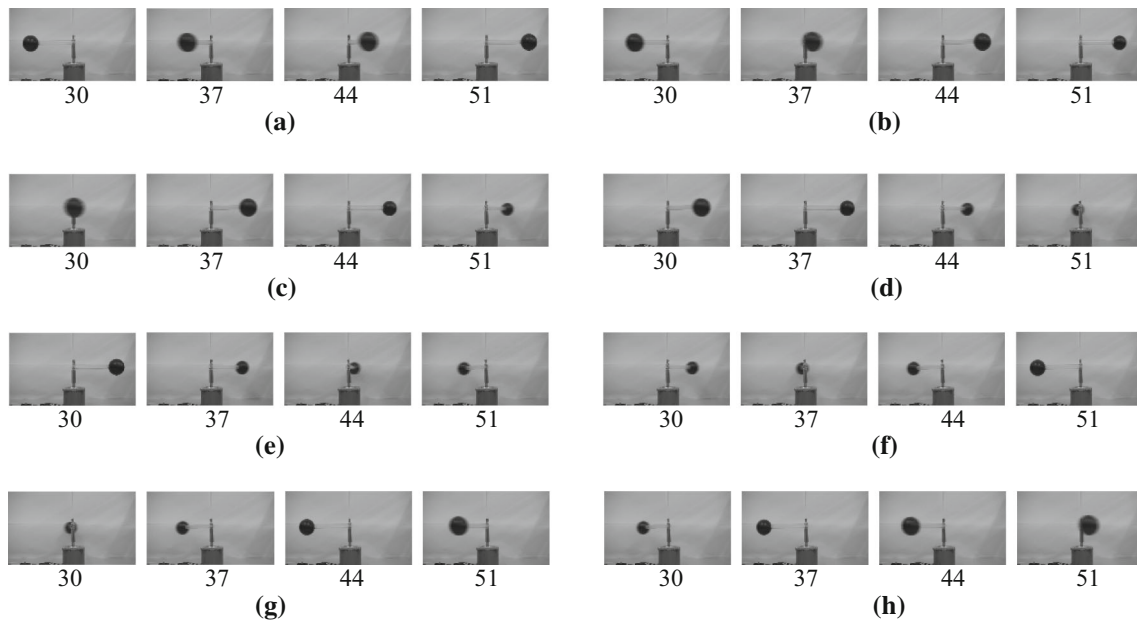
Fig. 19 Example frames of depth rotation with different starting points. Each video sequence is illustrated only by picking up four frames; the frame number is indicated under each image. It should be noted that in each video sequence, the first illustrative image (i.e., frame 30) represents the starting point of depth rotation of the ball
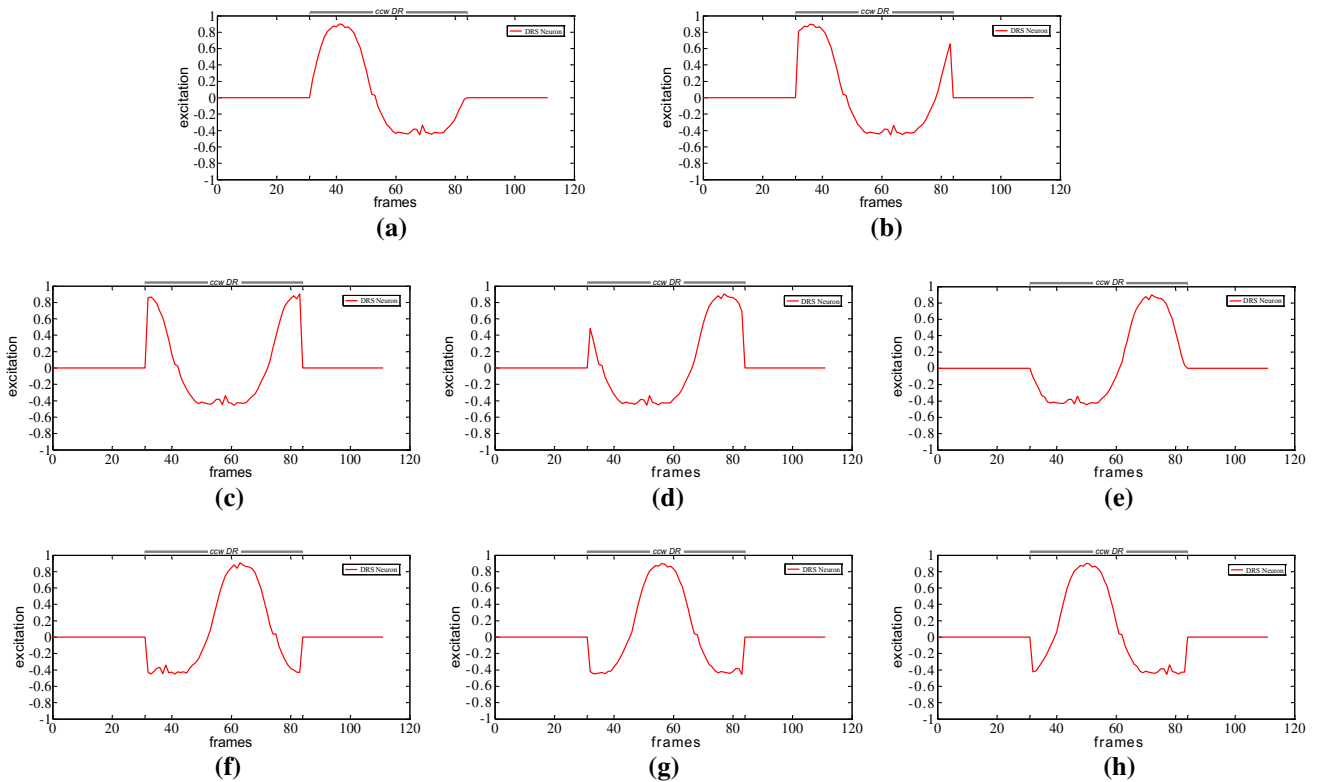

Fig. 20 Output curves of DRPNN. Each subgraph is acquired by DRPNN through the corresponding video sequence in Fig. 19

rotation for example to illustrate the schematic of the sight axis deviation by Fig. 21.

At the begin of sampling a video sequence, the camera's sight axis coincides with the Z axis. Then, the sagittal deviation angle between the axis and the Z axis increases 15° step by step till that the axis approaches the Y axis. When the axis deviation reaches over 75°, we turn it into 85° for the final video sequence. The acquired video
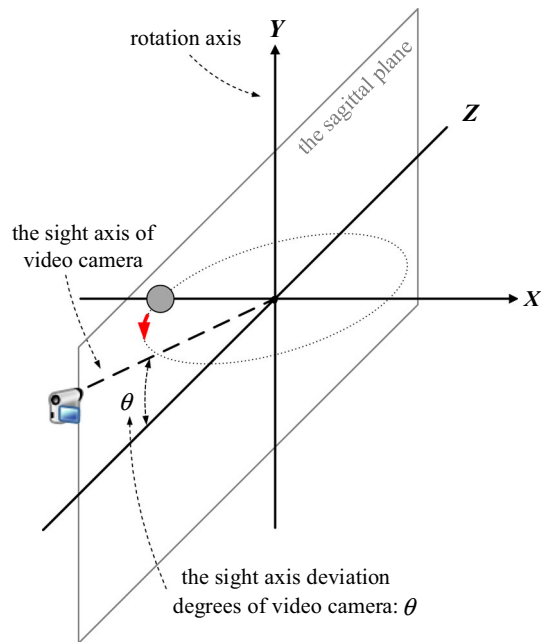
Fig. 21 Schematic illustration of the sight axis deviation

sequences are given in Fig. 22 below. In these video sequences, the motion patterns of the object are similar to that of the ball shown in Fig. 9a except that its rotation angular velocity is 3.55 rad/s.

As associated to Fig. 22, the output curves of DRPNN are plotted by Fig. 23. By these curves, we find that DRPNN can perceive the rotational motion of the object. When the sight axis deviation angle is small up to 30°, the output curves of DRPNN are quasi-sinusoidal (Fig. 23a, b). However, when the sight axis deviation angle increases gradually, the output curve of DRPNN gradually changes into a square waveform curve (Fig. 23c–e). When the sight axis is almost parallel to the rotation axis of the ball, the output curve is close to a square wave (Fig. 23f). This is

because that, when the sight axis deviates from the Z axis, the projection of a rotating object changes from a swing line to an ellipse, and the perceived depth rotation cues are gradually reduced. In the extreme case where the sight axis is perpendicular to the rotating plane of the rotating object, the projection of the object will form a circle, which leads to that the depth rotation pattern of the object fully disappears in the field of view. All the test results indicate that the spatio-temporal change of depth rotation is quasi-sinusoidal if the sight axis of the camera is only with small perturbation.

5.5 Comparative analysis

As far as we know, there is no appropriate computational model for depth rotation perception in the literature up to date. Here, we can only take the three recent motion perception neural networks to participate in comparative analysis, i.e., Beardsley's neural network [58], LGMD model [55] and RMPNN [10]. To compare DRPNN with the three mentioned neural networks, four depth rotation patterns in Fig. 9 are employed, i.e., ccw, cw, ccw, and cw depth rotation. More details can be found in Sect. 5.2. We here emphasize that in Fig. 9a, b, the ball makes ccw and cw depth motion from frame 32 to 60 and frame 34 to 61, respectively; in Fig. 9c, d, so does the rectangle from frame 31 to 52 and from frame 36 to 57, respectively.

A. Beardsley's neural network

Beardsley's model is a conventional three-layer back-propagation neural network. The input stimuli are the motion patterns represented by idealized optic flow, while the output layer includes sixteen MSTd units which prefer different types of motion patterns. Its input optic flow, caused by the depth rotation in each video sequence is acquired by the Lucas-Kanade method. Afterward, the
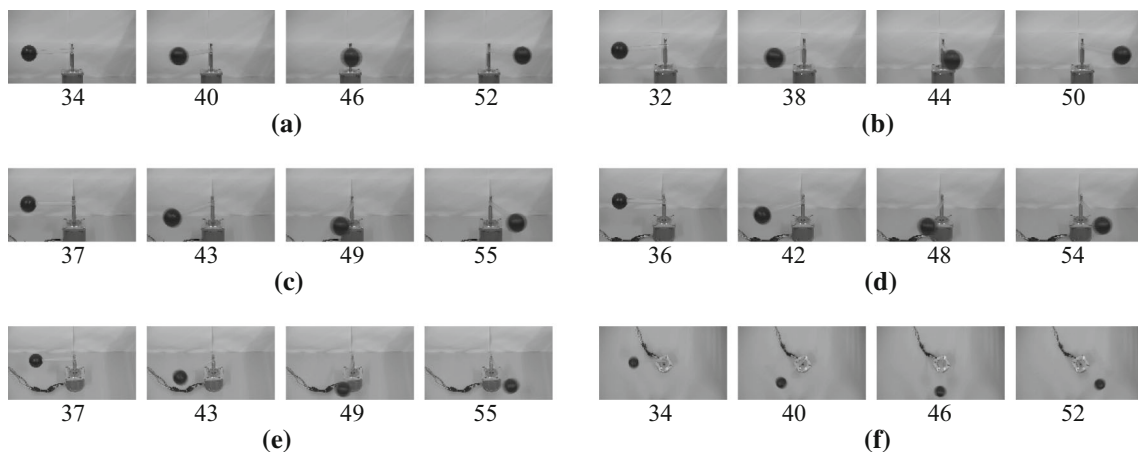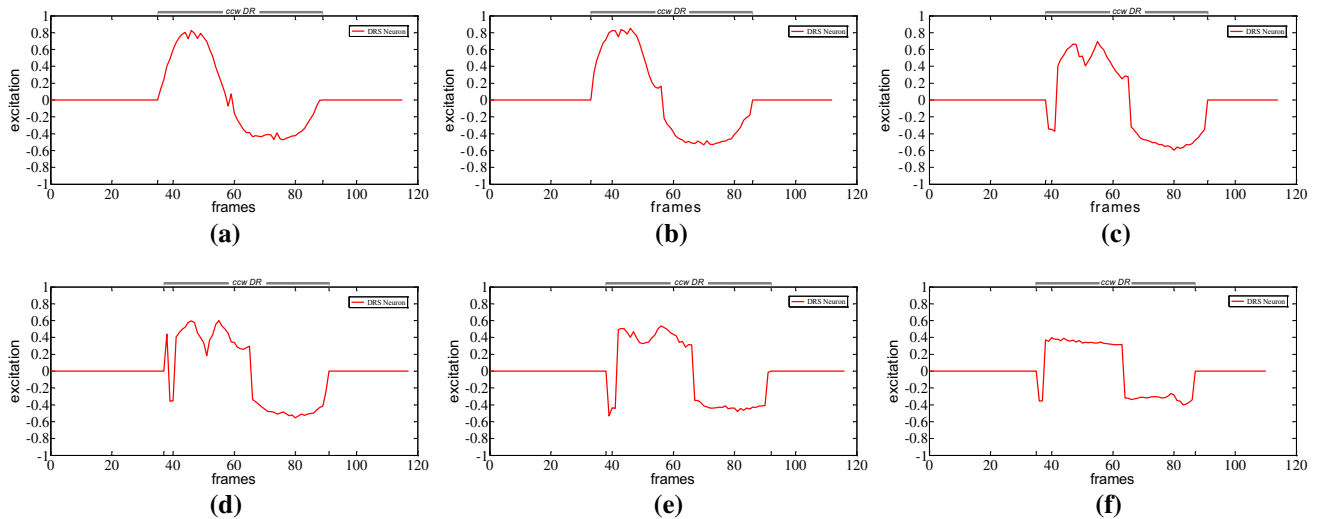

Fig. 22 Example frames for the sight axis deviation under ccw rotation on the horizontal plane. Each video sequence is illustrated only by picking up four frames; the frame number is indicated under each image

Fig. 23 Output curves of DRPNN. Each subgraph is acquired by DRPNN through the corresponding video sequence in Fig. 22
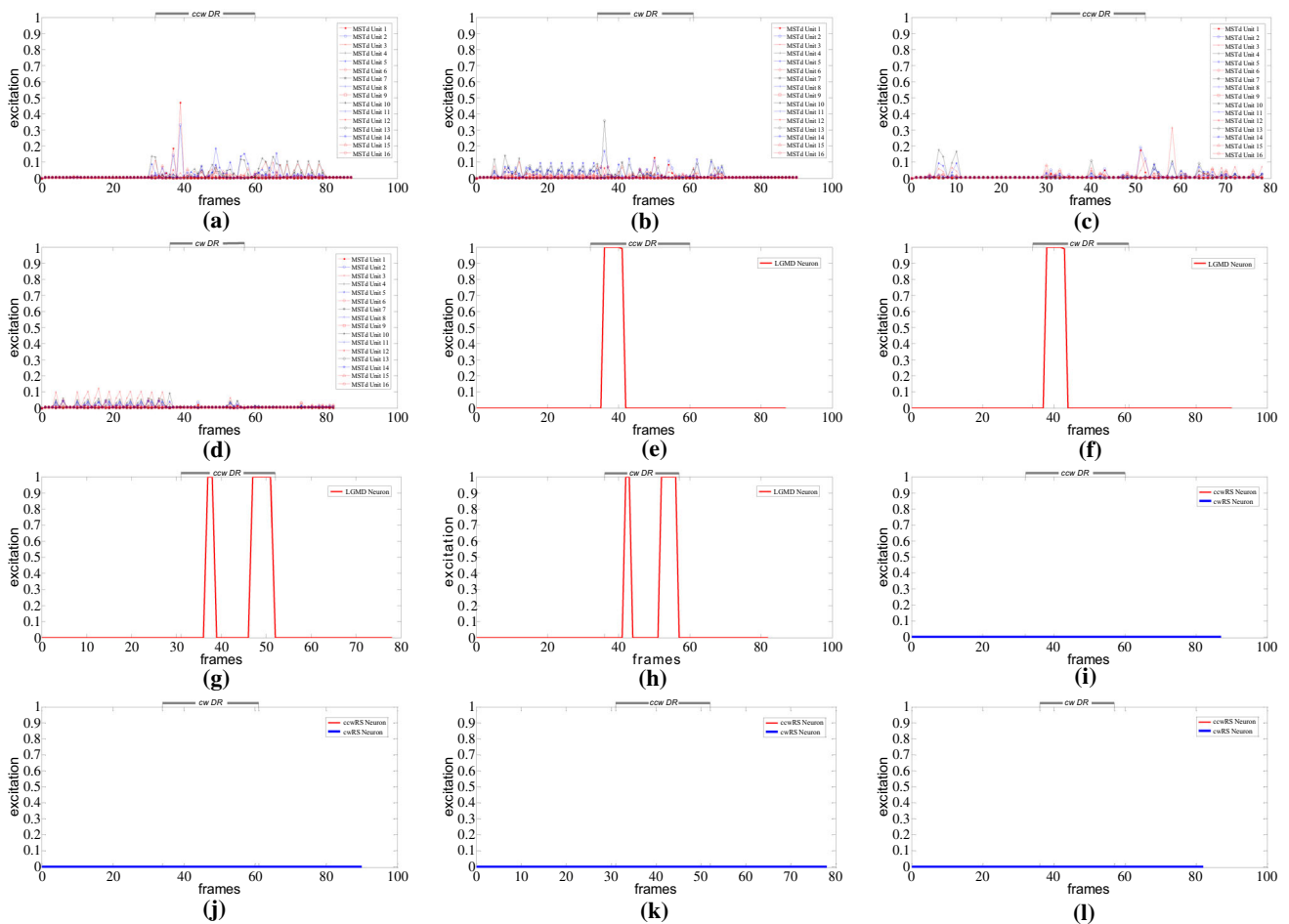

Fig. 24 Output curves of the three compared neural networks: **a–d** Beardsley's neural network, **e–h** LGMD model, and **i–l** RMPNN

output curves of the MSTd units are displayed in Fig. 24a–d in terms of Fig. 9. By comparing Fig. 10 with Fig. 24a–d, Beardsley's neural network cannot respond to depth

rotation in the real scene test, as the training samples of the network are required to be idealized [58] and the gained weight matrices only suit to those idealized optic flow

samples in virtual environments. Therefore, Beardsley's model fails to perceive depth rotation in real scenes.

B. LGMD model

LGMD mainly consists of four neural layers and one neuron, in which input stimuli are the frames extracted from video sequences. Based on the above four video sequences in Fig. 9, LGMD generates four output excitation curves as in Fig. 24e–h. Such curves can conclude that LGMD cannot correctly detect the spatio-temporal energy change of depth rotation. We here take the video sequence in Fig. 9c for example to analyze the performance of LGMD against depth rotation. The excitation curve in Fig. 24c indicates that, when the ball rotates clockwise on the horizontal plane, LGMD has no response with a long time, but it can become excitatory and discharge membrane potentials, as any depth rotation contains approaching motion component [24, 28, 32]. Therefore, depth rotation can also make LGMD keep excitatory when an object approaches toward the video camera. In Fig. 9c, when the rectangle passes through the two segments, i.e., P_2 to P_3, and P_8 to P_1 (Fig. 18), it will trigger LGMD to generate collision alarming, which illustrates that LGMD is effective for collision detection.

3 RMPNN

RMPNN includes two types of sub-networks. One is ccwRMPNN which responds to the ccw rotational motion, and the other is cwRMPNN which reacts to the cw rotational motion. Its output is rotation sensitive neuron's preference to rotational motion on the fronto-parallel plane. The output curves are shown in Fig. 24i–l, relying upon the visual frame stimuli extracted from video sequences in Fig. 9. The curves indicate that both ccwRMPNN and cwRMPNN have no response to any depth rotation, as RMPNN identifies its preferred motion patterns through detecting the continuous change in the translational motion direction on the fronto-parallel plane [10]. Particularly, the left/right translational and approaching/receding motion cues, generated by the depth rotation on the horizontal plane cannot make RMPNN become excitatory. Therefore, RMPNN could not respond to depth rotation in the field of view.

Summarily, compared by DRPNN, the above three types of motion perception neural networks exhibit their intrinsic characteristics and also expose their defects in solving the problem of depth rotation detection. Based on the above comparative experiments, we can draw some conclusions: (1) as a specially designed novel computational model, DRPNN is suitable for detecting the spatio-temporal energy change of depth rotation in real scenes; (2) Beardsley's neural network is designed to recognize motion patterns represented by idealized optic flow, while it is hard when detecting the unknown patterns of depth rotation in non-ideal scenes; (3) as a specific collision detection neural network, LGMD can detect the approaching motion included in depth rotation, whereas it fails to perceive the spatio-temporal energy change of depth rotation; and (4) even though RMPNN as a specific neural network for rotational motion can recognize rotational motion patterns, it cannot detect the depth rotation of an object.

5.6 Discussion

In the above sections, the presented DRPNN has been sufficiently examined by several types of depth rotation video sequences under various conditions. All of these experiments have verified the reliable ability of DRPNN in detecting depth rotation motion. The experimental results indicated that the properties of DRPNN coincide with most of the main functional properties of DRS neurons [7, 9, 11, 12], e.g., depth rotation detection, motion direction selection, position invariance, sensitivity on rotation speed and starting point. Also, DRPNN is compared with the three state-of-the-art motion perception models, by which the comparative results have demonstrated that DRPNN is effective for depth rotation object detection.

Although DRPNN can simply simulate some properties of visual information processing in biological vision systems, it cannot avoid two common defects in the field of artificial visual neural network: (1) when an object is in depth rotation at an extremely slow or fast rotation angular velocity, the extraction of motion cues is difficult, and hence DRPNN cannot correctly capture the motion characteristics of the moving object, and (2) DRPNN can only recognize the depth rotation of a moving object on some certain planes including horizontal, left diagonal, sagittal, and right diagonal planes, which means that it might not work well to those on other rotating planes.

On the other hand, the hypothesis in projective geometry, proposed by Johansson et al.[29] holds the viewpoint that, when a line segment is in depth rotation, the change of its projected length is similar to a sinusoidal curve, which has been confirmed by our experiments. Thus, DRPNN is an alternative model for depth rotation object detection.

6 Conclusion

Although many computational models have been developed for motion pattern detection, it is still rare to study how to detect the depth rotation pattern of an object. Thus, this work aims to develop a novel depth rotation perception neural network (DRPNN) in order to deal with the hard problem of depth rotation perception in computer vision. Since one such work reconciles with the related studies on visual information processing in biological vision systems,

it can bring some lights toward the building of artificial vision systems which integrate visual neurophysiologic findings with computer vision technologies for such motion detection tasks as visual motion perception, visual motion pattern recognition, intelligent video surveillance, autonomous robot and so on.

Inspired by the internal structure of the mammalian's retina and the functional properties of depth rotation sensitive neurons in neurophysiology, DRPNN is suggested to not only simulate the framework of hierarchical visual information processing, but also detect the spatio-temporal energy change of depth rotation of an object in the field of view. Comprehensive experiments are used to examine DRPNN's performance characteristics. The experimental results can draw three points: (1) DRPNN can recognize the depth rotation motion pattern of an object, (2) DRPNN is robust to the object's rotating plane and motion position in the field of view, and (3) DRPNN is sensitive to the object's rotation angular velocity and starting point in rotation trajectory as well as the camera's sight axis deviation. All these intrinsic properties of DRPNN can simply explain some functional properties of depth rotation sensitive neurons in the posterior parietal association cortex of primates. As the first bio-inspired computational model for depth rotation perception, this research is a significant step toward both intensively understanding visual information processing mechanisms in biological vision systems and probing into bio-inspired computational models for depth rotation object detection. In the future, DRPNN can be extended by integrating other bio-inspired visual neural networks for complex visual detection tasks. It can be used to construct an artificial vision system for engineering applications, e.g., gear/propeller rotation fault monitoring.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Yan C, Xie H, Yang D et al (2018) Supervised hash coding with deep neural network for environment perception of intelligent vehicles. IEEE Trans Intell Transp Syst 19:284–295

2. Vlasits A, Baden T (2019) Motion vision: a new mechanism in the mammalian retina. Curr Biol 29:R933–R935

3. Koenderink JJ, van Doorn AJ (1976) Local structure of movement parallax of the plane. J Opt Soc Am 66:717–723

4. Verri A, Girosi F, Torre V (1990) Differential techniques for optical flow. J Opt Soc Am A 7:912–922

5. Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 49:1127–1147

6. Rind FC, Simmons PJ (1999) Seeing what is coming: building collision-sensitive neurones. Trends Neurosci 22:215–220

7. Saito H, Yukie M, Tanaka K et al (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. J Neurosci 6:145–157

8. Sakata H, Shibutani H, Kawano K, Harrington TL (1985) Neural mechanisms of space vision in the parietal association cortex of the monkey. Vis Res 25:453–463

9. Sakata H, Shibutani H, Ito Y, Tsurugai K (1986) Parietal cortical neurons responding to rotary movement of visual stimulus in space. Exp Brain Res 61:658–663

10. Hu B, Yue S, Zhang Z (2017) A rotational motion perception neural network based on asymmetric spatiotemporal visual information processing. IEEE Trans Neural Netw Learn Syst 28:2803–2821

11. Sakata H, Shibutani H, Ito Y et al (1994) Functional properties of rotation-sensitive neurons in the posterior parietal association cortex of the monkey. Exp Brain Res 101:183–202

12. Sakata H, Taira M, Kusunoki M et al (1997) The parietal association cortex in depth perception and visual control of hand action. Trends Neurosci 20:350–357

13. Wang H, Peng J, Zheng X, Yue S (2020) A robust visual system for small target motion detection against cluttered moving backgrounds. IEEE Trans Neural Networks Learn Syst 31:839–853

14. Shojaei K (2019) Three-dimensional neural network tracking control of a moving target by underactuated autonomous underwater vehicles. Neural Comput Appl 31:509–521

15. Li L, Zhang Z, Lu J (2021) Artificial fly visual joint perception neural network inspired by multiple-regional collision detection. Neural Netw 135:13–28

16. Fu Q, Hu C, Peng J et al (2020) A robust collision perception visual neural network with specific selectivity to darker objects. IEEE Trans Cybern 50:5074–5088

17. Maheshan MS, Harish BS, Nagadarshan N (2019) A convolution neural network engine for sclera recognition. Int J Interact Multimed Artif Intell 6:78–83

18. Liu D, Bellotto N, Yue S (2020) Deep spiking neural network for video-based disguise face recognition based on dynamic facial movements. IEEE Trans Neural Netw Learn Syst 31:1843–1855

19. Jha S, Dey A, Kumar R, Kumar-Solanki V (2019) A novel approach on visual question answering by parameter prediction using faster region based convolutional neural network. Int J Interact Multimed Artif Intell 5:30–37

20. Hu B, Zhang Z, Li L (2019) LGMD-based visual neural network for detecting crowd escape behavior. In: Proceedings 2018 5th IEEE international conference cloud computing intelligent systems, CCIS 2018, vol 6, pp 772–778

21. Chen J, Su W, Wang Z (2020) Crowd counting with crowd attention convolutional neural network. Neurocomputing 382:210–220

22. Braunstein ML (1972) Perception of rotation in depth: a process model. Psychol Rev 79:510–524

23. Hershberger WA, Stewart MR, Laughlin NK (1976) Conflicting motion perspective simulating simultaneous clockwise and

counterclockwise rotation in depth. J Exp Psychol Hum Percept Perform 2:174–178

24. Braunstein ML (1984) Perception of rotation in depth: the psychophysical evidence. ACM SIGGRAPH Comput Graph 18:25–26

25. Shulman GL (1991) Attentional modulation of mechanisms that analyze rotation in depth. J Exp Psychol Hum Percept Perform 17:726–737

26. Braunstein ML (1976) Depth perception through motion. Academic Press, London

27. Petersik JT (1980) Rotation judgments and depth judgments: separate or dependent processes? Percept Psychophys 27:588–590

28. Andersen GJ, Braunstein ML (1983) Dynamic occlusion in the perception of rotation in depth. Percept Psychophys 34:356–362

29. Johansson G, Jansson G (1968) Perceived rotary motion from changes in a straight line. Percept Psychophys 4:165–170

30. Carpenter DL, Dugan MP (1983) Motion parallax information for direction of rotation in depth: order and direction components. Perception 12:559–569

31. Miles FA (1998) The neural processing of 3-D visual information: evidence from eye movements. Eur J Neurosci 10:811–822

32. Schaafsma SJ, Duysens J, Gielen CCAM (1997) Responses in ventral intraparietal area of awake macaque monkey to optic flow patterns corresponding to rotation of planes in depth can be explained by translation and expansion effects. Vis Neurosci 14:633–646

33. Simmons PJ, Rind FC, Santer RD (2010) Escapes with and without preparation: the neuroethology of visual startle in locusts. J Insect Physiol 56:876–883

34. Rind FC, Bramwell DI (1996) Neural network based on the input organization of an identified neuron signaling impending collision. J Neurophysiol 75:967–985

35. Yue S, Rind FC (2006) Collision detection in complex dynamic scenes using an LGMD-based visual neural network with feature enhancement. IEEE Trans Neural Netw 17:705–716

36. Yue S, Rind FC (2006) Visual motion pattern extraction and fusion for collision detection in complex dynamic scenes. Comput Vis Image Underst 104:48–60

37. Yue S, Rind FC (2013) Postsynaptic organisations of directional selective visual neural networks for collision detection. Neurocomputing 103:50–62

38. Gabriel JP, Trivedi CA, Maurer CM et al (2012) Layer-specific targeting of direction-selective neurons in the zebrafish optic tectum. Neuron 76:1147–1160

39. Bereshpolova Y, Stoelzel CR, Su C et al (2019) Activation of a visual cortical column by a directionally selective thalamocortical neuron. Cell Rep 27:3733–3740

40. Fried SI, Münch TA, Werblin FS (2002) Mechanisms and circuitry underlying directional selectivity in the retina. Nature 420:411–414

41. Huang X, Rangel M, Briggman KL, Wei W (2019) Neural mechanisms of contextual modulation in the retinal direction selective circuit. Nat Commun 10:1–15

42. Fu Q, Yue S (2017) Modeling direction selective visual neural network with ON and OFF pathways for extracting motion cues from cluttered background. In: 2017 International joint conference on neural networks (IJCNN) IEEE 831–838

43. Fu Q, Yue S (2020) Modelling Drosophila motion vision pathways for decoding the direction of translating objects against cluttered moving backgrounds. Biol Cybern 114:443–460

44. Wei W (2018) Neural mechanisms of motion processing in the mammalian retina. Annu Rev Vis Sci 4:165–192

45. Vlasits AL, Euler T, Franke K (2019) Function first: classifying cell types and circuits of the retina. Curr Opin Neurobiol 56:8–15

46. Morrone MC, Burr DC, Vaina LM (1995) Two stages of visual processing for radial and circular motion. Nature 376:507–509

47. Fu Q, Wang H, Hu C, Yue S (2019) Towards computational models and applications of insect visual systems for motion perception: a review. Artif Life 25:263–311

48. Grünert U, Martin PR (2020) Cell types and cell circuits in human and non-human primate retina. Prog Retin Eye Res 78:1–33

49. Field GD, Rieke F (2002) Nonlinear signal transfer from mouse rods to bipolar cells and implications for visual sensitivity. Neuron 34:773–785

50. Gollisch T, Meister M (2010) Eye smarter than scientists believed: neural computations in circuits of the retina. Neuron 65:150–164

51. Yang X, Wu SM (1991) Feedforward lateral inhibition in retinal bipolar cells: Input-output relation of the horizontal cell-depolarizing bipolar cell synapse. Proc Natl Acad Sci 88:3310–3313

52. Thoreson WB, Mangel SC (2012) Lateral interactions in the outer retina. Prog Retin Eye Res 31:407–441

53. Rind FC, Wernitznig S, Pölt P et al (2016) Two identified looming detectors in the locust: ubiquitous lateral connections among their inputs contribute to selective responses to looming objects. Sci Rep 6:1–16

54. Hu B, Zhang Z (2018) Bio-plausible visual neural network for spatio-temporally spiral motion perception. Neurocomputing 310:96–114

55. Yue S, Rind FC (2013) Redundant neural vision systems-competing for collision recognition roles. IEEE Trans Auton Ment Dev 5:173–186

56. Albright TD, Desimone R, Gross CG (1984) Columnar organization of directionally selective cells in visual area MT of the macaque. J Neurophysiol 51:16–31

57. Schneider M, Kemper VG, Emmerling TC et al (2019) Columnar clusters in the human motion complex reflect consciously perceived motion axis. Proc Natl Acad Sci U S A 116:5096–5101

58. Beardsley SA, Ward RL, Vaina LM (2003) A neural network model of spiral-planar motion tuning in MSTd. Vision Res 43:577–595