Neural Networks 135 (2021) 13–28

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Artificial fly visual joint perception neural network inspired by multiple-regional collision detection

Lun Li, Zhuhong Zhang *, Jiaxuan Lu

College of Big Data and Information Engineering, Guizhou University, Guizhou Provincial Characteristic Key Laboratory of System Optimization and Scientific Computing, Guiyang, Guizhou 550025, PR China

ARTICLE INFO

Article history:
Received 26 March 2020
Received in revised form 12 November 2020
Accepted 30 November 2020
Available online 5 December 2020

Keywords:
Fly visual neurophysiology
Fly's vision system
Feedforward neural networks
Multi-regional collision detection
Perceptual region division

ABSTRACT

The biological visual system includes multiple types of motion sensitive neurons which preferentially respond to specific perceptual regions. However, it still keeps open how to borrow such neurons to construct bio-inspired computational models for multiple-regional collision detection. To fill this gap, this work proposes a visual joint perception neural network with two subnetworks — presynaptic and postsynaptic neural networks, inspired by the *preferential* perception characteristics of three horizontal and vertical motion sensitive neurons. Related to the neural network and three hazard detection mechanisms, an artificial fly visual synthesized collision detection model for multiple-regional collision detection is *originally* developed to monitor possible danger occurrence in the case where one or more moving objects appear in the whole field of view. The experiments can clearly draw two conclusions: (i) the acquired neural network can effectively display the characteristics of visual movement, and (ii) the collision detection model, which outperforms the compared models, can effectively perform multiple-regional collision detection at a high success rate, and only takes about 0.24s to complete the process of collision detection for each virtual or actual image frame with resolution 110×60.

1. Introduction

As an ancient but active research branch, collision detection frequently involves many engineering-related technology fields such as car collision, navigation, aerospace, ocean and so on. It has comprehensively been studied over the past decades, as some collision detection sensors such as laser, radar and ultrasonic sensors can successfully perform information processing and also output specific information, e.g. location, velocity, distance, etc. Such types of sensors, however, bring about economically expensive cost in engineering, in order to acquire highly precise detection performance. Fortunately, bio-inspired visual sensors as a visual information-processing system is a new type of processor which can capture any visual information, and in particular those desired objects in the field of view. This will be an important and potential research branch, since bio-visual neural mechanisms as a useful tool of collision detection can be simply modeled to formulate how one moving object approaches another one.

As a valuable model organism, the fly vision system can be adopted to study the neural basis of visual behavior, due to the merits of high speed motion, free navigation and especially efficient colliding avoidance in complex scenes (Maisak, Haag, Ammer, et al., 2013; Schneider, Murali, Taylor, et al., 2018). When an object moves in the field of view, flies can not only perceive and detect spatiotemporal intensity patterns, but also avoid possible collision as soon. Particularly, their visual neural systems can be viewed as a visual sensor to process those received visual clues. This is a nice bio-inspiration to explore artificial visual systems for object tracking (Aptekar, Shoemaker, & Frye, 2012; Fox & Frye, 2014; Missler & Kamangar, 1995), navigation control (Kim & Dickinson, 2017; Zufferey & Floreano, 2006) and collision detection (Pitonakova & Bullock, 2020; Zhang, Yue, & Zhang, 2015). However, whereas the flies can exhibit excellent behaviors, can their visual neural properties be simulated to construct bio-inspired computational models? Particularly, can their direction-selective neurons such as horizontal neurons H1 and H3, vertical neuron V1 and the lobula plate tangential cell (LPTC) be synthesized to develop an artificial fly visual joint perception neural network for multiple-regional collision detection? This open issue will be discussed in the current research.

It is pointed out that the present motivation is to solve the problem of multiple-regional collision detection, based on bio-inspired fly visual neuronal information-processing mechanisms. The main *innovation* includes three points: (i) a specific parabolic curve is *originally* used to divide the medulla's direction detection region into three subregions, inspired by biological visual perception characteristics; (ii) an artificial fly visual joint perception neural network (AFVJPNN) is *originally* developed to detect the

* Corresponding author.
E-mail address: zhzhang@gzu.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.neunet.2020.11.018
0893-6080/

motion-directional changes of one or more moving objects in the *whole* field of view, motivated by the functional response properties of the fly vision system; and (iii) an artificial fly visual synthesized collision detection model is *originally* designed to perform multiple-regional collision detection.

It is also highlighted that AFVJPNN as a biologically-inspired visual neural network differs from any reported bio-visual collision detection models, due to different bio-inspirations and engineering applications. On one hand, AFVJPNN originates from the visual response properties of three direction-selective neurons (i.e. H1, V1 and H3 neurons) discovered by neurophysiologists in the fly vision system. Meanwhile, a visual synthesized collision detection model is designed to execute multiple-regional collision detection *for the first time*. On the other hand, the reported bio-visual models *can only* execute *single-regional* collision detection. Especially, AFVJPNN is different from our previous artificial fly visual neural network (AFVNN, Zhang et al., 2015) and those developed by Yue's group (Hu, Arvin, Xiong, et al., 2017; Yue & Rind, 2006a, 2006b). Whereas both AFVNN and AFVJPNN originate from the same fly visual neural mechanism, their design inspirations are different. The former one *can only* react to a *single* moving object in the *front* of the field of view, whereas the latter one can simultaneously respond to multiple moving objects and detect the change of visual motion in the *whole* field of view, and especially it can capture possible object information in the *left and right* visual regions. Additionally, the reported and representative visual neural networks, acquired by Yue's group are developed to simulate the functional characteristics of locust-based lobula giant movement detector (LGMD), inspired by locust visual neural mechanisms. Such networks with multiple presynaptic direction-selective neural networks can detect *single-regional* motion change for a single moving object, but they cannot execute multiple-regional collision detection.

The rest of this paper is organized as follows. Section 2 presents an overview of related work. Some fly visual neurophysiologic theories and a related neurocomputational model are given in Section 3. AFVJPNN is exhaustively designed in Section 4. Section 5 formulates an artificial fly visual synthesized collision detection model. The whole experimental study is given in Section 6. Finally, some conclusions and open issues appear in Section 7.

2. Survey on related work

Studies on the fly visual system mainly concentrate on both anatomizing its visual neuronal structure and developing visual neurocomputational models related to specific engineering applications (Dalgaty, Vianello, De Salvo, et al., 2018; Ma & Krings, 2009; Taherkhani, Belatreche, Li, et al., 2020). Some prominent achievements have initially involved such topics as motion detection (Eichner, Joesch, Schnell, et al., 2011), object tracking (Sun, Liu, & Guo, 1999), collision detection (Blanchard, Rind, Verschure, et al., 2000; Yue & Rind, 2013) and autonomous obstacle avoidance (Pant & Higgins, 2012).

2.1. Motion sensitive neuron model

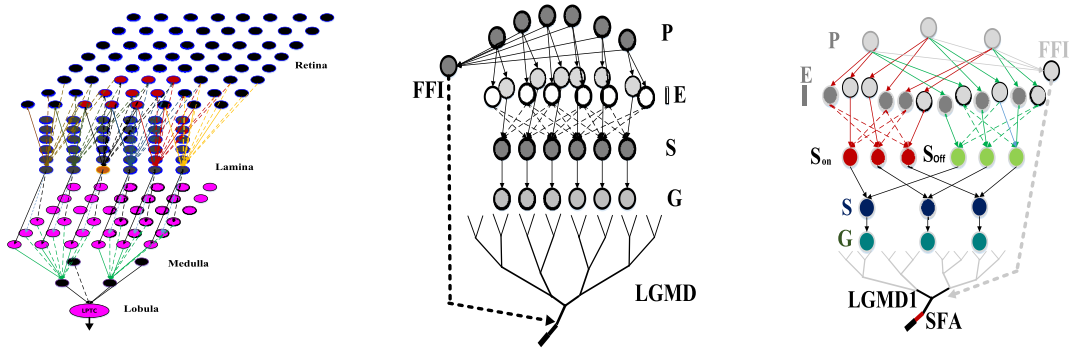
Since Bishop and Keehn (1966) first discovered fly's motion sensitive neurons in 1966, electro- and neuro-scientists (Mauss & Borst, 2016) have intensively studied the interactive relation between such neurons as the large monopolar cells and lobula plate tangential cells (LPTCs). Especially, Hausen and Eckert (Eckert, 1980; Hausen, 1976) reported eight horizontal and vertical neurons in the fly's optic lobe structure that performed motion-directional detection, e.g. horizontal neurons H1, H2 and H3 as well as vertical neurons V1 and V2. Subsequently, some motion detectors in the lamina's neural structure were found by several research groups. Tuthill, Nern, Holtz, et al. (2013) discovered twelve neurons (L1–L12) that performed depolarization and hyper-polarization for visual motion recordings, and meanwhile developed an electro-physiologic circuit model; Clark, Bursztyn, Horowitz, et al. (2011) demonstrated that L1 and L2 could strongly respond to the edges of light and dark objects, and meanwhile designed a differentially weighted edge selectivity model to detect rotational motion. Additionally, fly's lobula complex as a central organism in the visual neural system has been initially discussed (Longden, Wicklein, Hardcastle, et al., 2017). Some findings indicate that its visual clues can be transmitted to the brain nervous system in terms of directional membrane potentials produced by visual neural layers of retina, lamina, medulla and lobula. Keles and Frye (2017) identified a lobula columnar neuron (LC11) that could accurately perform collision detection in the fly's optic lobe. However, although all these achievements are valuable in electro- or neuro-physiology, it is still open how to design computational models to describe the relation between the input and the output for those motor-stimulated neurons.

2.2. Neurocomputational model

The fly visual system as a biological organism plays an important role in exploring neurocomputational models for neural circuit analysis, motion detection, object tracking, collision detection and collision avoidance.

Motion detection is one of the oldest but active topics in the field of computer vision. Herein, fly's vision system as an excellent bio-organism can be used to construct visual perceptron models for detecting whether there exists a moving object in the field of view (Clifford & Ibbotson, 2002). Hassenstein and Reichardt (1956) developed an elementary motion detector (EMD) to detect the change of motion of an object in the horizontal direction. Eichner et al. (2011) suggested a four-channel detector and a two-quadrant motion detector to analyze the process of visual transmission by virtue of fly's visual internal neural structure. In order to estimate the flying speed of flies in a cluttered environment, Wang, Dewell, Zhu et al. (2018) and Wang, Peng, Baxter et al. (2018) announced an angular velocity detector model of which each unit included three neighboring photoreceptors and two angular velocity detectors with different sampling rates. Wang, Peng, and Yue (2017) designed a LPTC-based neural model to filter irrelevant motion signals in clutter background, depending on the estimations of background motion direction, EMD and a two-quadrant motion detector.

Object tracking is another active topic in computer vision, which mainly focuses on exploring efficient computational models to estimate real-time information on position, velocity and distance for a moving object. Therein, Missler and Kamangar (1995), for the first time, developed an artificial fly visual neural network with four neural layers by means of the internal structure of fly's visual lobe. The neural network can only estimate the position and velocity of a small object in the black or white environment. Sun et al. (1999) designed a neurocomputational model related to EMD, an oscillatory neural network and a selective attention model. They claimed that such a model could explain some properties of the primate visual cortex and also suit to tracking any object in the temporal and spatial domains. Nasir and Mat (2019) designed a three-dimensional automatic visual tracking device with strong robustness. Fu, Bellotto, Hu et al. (2018) developed a hybrid fly visual neural network by means of the functional properties of two bio-plausible visual neural channels, in order to analyze object's moving performance and solve the problems of motion tracking and navigation by an autonomous micro biological robot.



(a) AFVNN (Zhang et al., 2015) (b) LGMD (Yue et al., 2006) (c) LGMD1 (Fu et al., 2018)

Fig. 1. (a) The fly visual neural network consists of four layers (Retina, Lamina, Medulla, Lobula) and a neuron (LPTC); (b) the locust visual neural network includes five layers (P, E, I, S, G) and two cells (FFI, LGMD); (c) the visual neural network is an improved version with six layers (P, E, I, S_{on}/S_{off}, S, G) and three cells (FFI, LGMD1, SFA).

Collision detection as a challenging topic involves how to explore a computational model capable of detecting imminent objects and deciding when a collision alarm will be transmitted. Visual neurophysiology studies have confirmed that, when an object approaches a fly, some visual neurons become excitatory and then output high membrane potentials in order to avoid possible collision. Based on such biological properties, several reported artificial visual models have been constructed to evaluate whether an alarm occurs in the *narrow* field of view. Our group developed AFVNN (Fig. 1(a), Zhang et al., 2015) to present the property of response of fly's LPTC, depending on its visual neural internal structure and EMD. The experimental results have validated that AFVNN is extremely sensitive to a small or large moving object, but cannot respond to objects located in edge regions. Additionally, as associated to other visual neural metaphors, some collision detection models have been continually reported (Oliva & Tomsic, 2014; Temizer, Donovan, Baier, et al., 2015). Liang, Morie, Suzuki, et al. (2007) developed a FPGA-based hybrid collision warning system related to hippocampal motion detection neurons, in which a special resistive-fuse network, the property of neuronal propagation in the hippocampus and also the fuzzy inference were used to execute coarse edge detection, danger evaluation and collision warning, respectively. Sun and Xiao (Sun & Frost, 1998; Xiao, Li, & Wang, 2006) designed three computational models based on several collision selective neurons in the round brains of pigeons, each of which measured a specific optical variant related to the expansion of the image of the imminent object. Particularly, Yue's group (Hu, Fu, Yue, et al., 2018; Yue & Rind, 2006a, 2006b, 2013) has developed a series of original or extended locust visual neural networks used for car collision detection, after intensively exploring the properties of the neurons of LGMD and LGMD1 in the locust's vision system. For example, based on the metaphor of neuronal directional inhibition, they, for the first time, developed an original and LGMD-based visual neural network with six-layer structure (Fig. 1(b), Yue & Rind, 2006a, 2006b), in order to detect object's collision in the visual front of the locust. One such model has been successfully applied to robot navigation. Furthermore, as associated to the fact that LGMD1 has a deep response to looming stimuli and possesses most of the same attributes as those in LGMD2 (Fu & Yue, 2015). Fu, Hu, Peng et al. (2018, 2019) proposed a collision detection model related to LGMD1 as in Fig. 1(c), in order to deal with the technical issue of selective collision detection of ground vehicles and bio-robots. Therein, the four on–off channels, based on a spike frequency adaptation mechanism are presented to exhibit visual cells' response preference with the help of a shunting mathematical model.

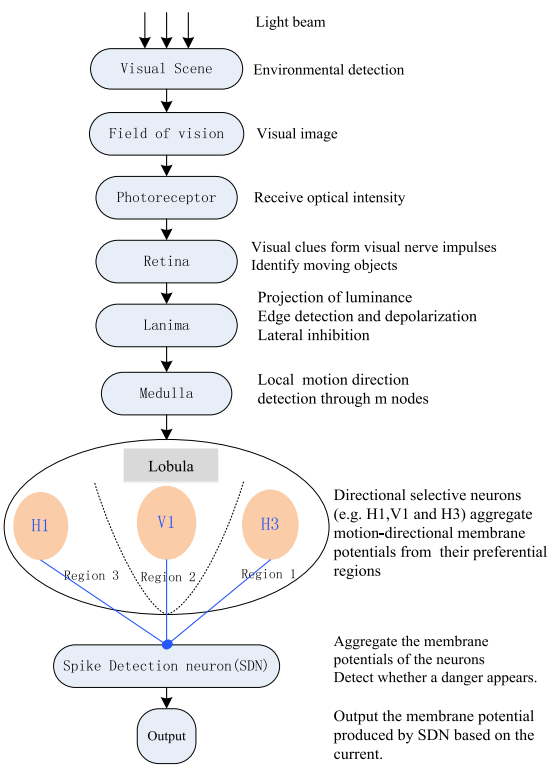
Collision avoidance as another challenging topic has been initially studied in the fields of visual neuroscience and computer vision. Lindemann, Kern, Van Hateren, et al. (2005) designed an experimental model which could extract translation velocity information on natural light flow stimulation in the inter-saccadic intervals, based on the unique dynamic characteristic of light flow perceived by the free flight of the fly. Moreover, the model can also quantitatively predict the collision response of complex dynamic stimuli encountered by flies in free flight. Hennig and Egelhaaf (2012) constructed a computational model composed of a spatial low-pass filter, a half-wave rectifier and EMD as well as multiple neural cells such as H1, HSE, etc. Pant and Higgins (2012) proposed a novel continuous-time neuronal model after executing experimental analysis on typical biological collision avoidance. Tammero and Dickinson (2002) developed a neurocomputational model based on the characteristics of fly's collision avoidance and landing response, relying upon a series of EMD-based detectors.

3. Fly's vision system and neurocomputational model

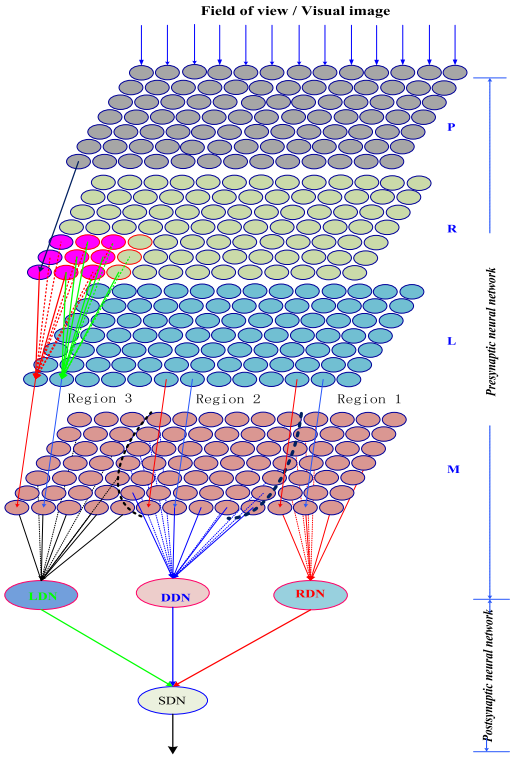
Some fundamental visual neurophysiological theories are introduced in Section 3.1. Subsequently, the simplified fly visual transmission and processing framework, based on the functional properties of visual neurons is displayed in Section 3.2, in order to acquire AFVJPNN and a valuable collision detection system.

3.1. Visual neurophysiology

The fly visual system consists of two compound eyes, each of which is composed of hexagonally arranged ommatidia (Borst, Haag, Reiff, et al., 2010; De Andresbragado & Sprecher, 2019). As an optic apparatus with a set of photoreceptors circularly arranged below a lens, each ommatidium hosts eight kinds of photoreceptors (R1–R8) surrounded by support and pigment cells (Ito, Shinomiya, Ito, et al., 2014). One such visual system collects external visual signals and processes them orderly by four neural layers, i.e. photoreceptor, retina, lamina and medulla (Fischbach & Dittrich, 1989). Photoreceptors receive external visual signals and perform figure-ground discrimination, while projecting their luminance intensities onto the next optic nerve lobe (Joly, Recher, Brombin, et al., 2016). In the retina layer, all ommatidia detect the presence of brightness in the field of view and perceive environmental changes. Each ommatidia continuously receives the brightness intensity of its counterpart in the photoreceptor layer and sends its membrane potential to the counterpart in the lamina after depolarization (Omoto, Keles, Nguyen, et al., 2017).



(a) Fly visual information-processing framework.



(b) Fly visual neural network framework.

Fig. 2. Fly visual neurocomputational model and the related neural network. (a) Once the field of view receives a visual scene, the fly's vision system performs visual transmission and processing through the five layers of photoreceptor, retina, lamina, medulla and lobula. In the lobula layer, each directional selective neuron only *reacts* to its preferential region in the medulla layer. (b) AFVJPNN simply simulates the mechanisms of visual response in Fig. 2(a), where P, R, L and M are the abbreviations of the photoreceptor, retina, lamina and medulla layers, respectively. Once a grayscale frame enters the P layer, it is processed orderly by each of the four layers. LDN, DDN and RDN generate motion-directional potentials through their preferential regions in the medulla layer, and transmit their potentials to SDN. SDN outputs a synthesized directional potential which expresses the status of each visual scene.

The lamina layer is composed of lamina units, each of which mainly consists of cartridge and on–off nodes. Each cartridge node receives the projections of luminance intensities around the counterpart in the retina layer after performing edge detection and depolarization. Each on–off node transmits its input signal to the motion detection circuit located in the medulla layer, after executing depolarization and hyperpolarization by means of an inhibition mechanism (Eichner et al., 2011). The medulla layer performs local and regional direction detection, and also sends each node's excitatory intensity to the lobula complex. The lobula plate in the complex involves in four neural layers, each of which corresponds to one of the left, right, down and up motion directions (Hausen, 1976; Buchner, Buchner, & Bülthoff, 1984), and also detects the motion-directional change of object(s) by a special LPTC (i.e. H1, H3, V1 and V2). Each of the LPTCs creates a visual synthesized membrane potential after gathering motion-directional potentials produced by medulla nodes, and subsequently sends its output to the central complex (Pfeiffer & Homberg, 2014) in order to form a visual information-processing circuit. The V2-cell is a unique heterolateral LPTC located between the layers of the horizontal and the vertical cells in the middle of the lobula plate. It generates spike activities and reacts to ipsilateral upward motion. Hereby, its strong response can be elicited by ipsilateral forward motion.

3.2. Visual transmission and neurocomputational model

The H1 neuron so-called left detection neuron (LDN) can selectively respond to ipsilateral horizontal forward motion, but inhibit reverse motion. However, the H3 neuron, i.e. a spike-transmitting and heterolateral motion sensitive neuron denoted here by right detection neuron (RDN), can only make specific response to ipsilateral backward motion. Additionally, as another spike-transmitting neuron, the V1 neuron so-called downward detection neuron (DDN) is strongly excited by ipsilateral downward motion. Each of the three neurons, together with the four neural layers mentioned above, constitutes the presynaptic neural network. When an object approaches a fly from top to bottom, from right to left, or from left to right, DDN, RDN or LDN will rapidly respond visual signals and produce a potential, and later transmit a spike through a common spike detection neuron (SDN), i.e. neuron V2. Since the lobula plate is almost covered by LPTCs' dendrites, a network of connections between LPTCs is formed to combine their responses and ultimately decides whether to trigger a spike signal or not (Cuntz, 2014). These neurophysiological findings motivate us to draw a simplified visual information-processing or neurocomputational model given by Fig. 2(a).

4. Joint perception neural network

This section first gives a strategy of region division in Section 4.1 used for deciding the preferential response subregion of the three neurons, i.e. LDN, DDN and RDN. Second, an artificial fly visual joint perception neural network (AFVJPNN) is designed in Sections 4.2 and 4.3.

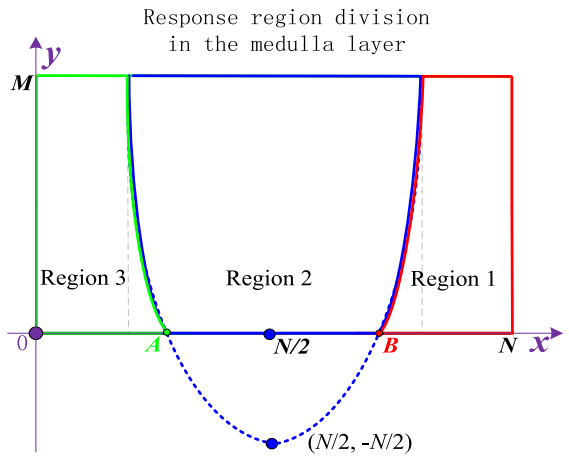


Fig. 3. Schematic illustration on response region division in the medulla layer. The upper half-plane is divided into three regions 1, 2 and 3 with the same angle of 60°.

Inspired by the process of visual hierarchical information-transmission and processing in the fly vision system, by an analogy with the properties of the fly's vision system in Fig. 2(a) AFVJPNN consists of presynaptic and postsynaptic neural networks, schematically illustrated by Fig. 2(b). The presynaptic neural network includes three neural subnetworks which share the three neural layers of photoreceptor, retina and lamina. Particularly, on the basis of the three common layers, each subnetwork only pays close attention to a specific region in the medulla layer. Their output neurons output membrane potentials by their distinctive response properties.

4.1. Preferential perception region

Biologically, a visual field usually involves wide and narrow visual fields. Based on bio-visual perceptual characteristics, the whole field of view can be divided into three subregions — left, front and right regions with angle 60° per subregion. The front of the field of view can coarsely match with a parabolic-like region. Correspondingly, by Fig. 3 a *rectangular* medulla region can be segmented into three subregions, since directional selective neurons only react to their preferential regions in the field of view. More precisely, in the pixel coordinate system the origin corresponds to the point in the Mth row and the 1st column; the x-axis is the horizontal direction line from left to right, while the y-axis represents the vertical direction line from down to upper; the parabolic curve with vertex $(N/2, -N/2)$ has two points $A(N/3,0)$ and $B(2N/3,0)$ on the x-axis.

4.2. Presynaptic neural network

As related to the preferential response characteristics of the three neurons as in Section 3.2, in the current work we appoint that LDN, DDN and RDN only respond to directional activities of medulla nodes from the left, front and right subregions by virtue of Fig. 3, respectively. By integrating Fig. 2(b) with Fig. 3, the presynaptic neural network consists of three fly visual feedforward neural networks (simply written as L-FVNN, D-FVNN and R-FVNN) which correspond to LDN, DDN and RDN, respectively. We take L-FVNN for example to explicitly formulate its design inspirations.

L-FVNN is used to detect the changes of local motion generated by one or more moving objects in the left visual region related to Region 3 in Fig. 3. It, schematically illustrated by Fig. 2(b) is composed of an output neuron (LDN) and four visual information-processing layers. The P-layer not only receives light luminance or grayscales of pixels in successive images, but also filters additional noise signals at the same pixel point by the conventional ViBe method (Barnich & Van Droogenbroeck, 2011), in order to extract some moving object(s) in the whole visual scene. In the R-layer, all the nodes receive the changes of brightness from the counterparts in the P-layer, while delivering their membrane potentials to the retinotopic counterparts in the L-layer after convolution operation. The L-layer decides each neural node's intensity of excitation or inhibition by an on–off mechanism. In the M-layer, only the nodes, which match with those in Region 3, produce motion-directional intensities, while other nodes keep no response. Finally, LDN gathers the local optical flows of such nodes, and produces its membrane potential.

(1) P-layer

The P-layer is formed of photoreceptors arranged in a matrix form with M rows and N columns, which receives the current image frame and only outputs its grayscale foreground after image preprocessing. Hereby, let V_f be a grayscale frame with size $M \times N$ at the moment of f, and $V_f(i, j)$ the luminance intensity the pixel (i, j) in V_f. Each photoreceptor corresponds to a pixel point in the image frame. Usually, the inter-frame difference (IFD) and ViBe methods (Barnich & Van Droogenbroeck, 2011; Weng, Huang, & Da, 2010) as two popular image preprocessing methods can be used to eliminate those background noises included in a series of image frames. Here, the ViBe method is chosen to create a $M \times N$ foreground matrix (i.e. L_f) in terms of V_f, in order to ensure that the photoreceptors only notice valuable visual signals. Accordingly, the excitatory intensity of photoreceptor (i, j), $P_f(i, j)$ with $1 \le i \le M$ and $1 \le j \le N$, is computed by

$$P_f(i,j) = \begin{cases} L_f(i,j), & \text{if } L_f(i,j) > 0, \\ 0, & \text{else}, \end{cases} \tag{1}$$

where $L_f(i, j)$ is denoted by the grayscale value at the point (i, j) in the matrix of L_f. Hence, based on the historical optical flow signals, each photoreceptor outputs its membrane potential by

$$\widetilde{P}_f(i,j) = P_f(i,j) + \sum_{l=1}^{n_c} \kappa_l P_{f-l}(i,j),\ 1 \le i \le M,\ 1 \le j \le N, \tag{2}$$

with maximal time step n_c, where κ_l is a persistence coefficient given by $\kappa_l = (1 + \exp(n_c \times l))^{-1}$. Eq. (1) indicates that only those valuable visual signals in V_f can be sent to the next layer (i.e. R-layer), which is helpful for capturing the performance feature of some moving object(s). Eq. (2) shows that, once one or more objects appear in the whole field of view, some photoreceptors can keep successive excitation or inhibition with a short time.

(2) R-layer

The R-layer consists of $M \times N$ retina nodes arranged in a matrix form. Each node first receives the membrane potentials of the 3×3 photoreceptors around its counterpart in the P-layer, and then outputs its projection by a spatiotemporal convolution kernel. In other words, the light intensity of node (i, j) in the layer, $R_{ij}(f)$, is computed by

$$R_{ij}(f) = \sum_{0 \le k,l \le 2} \widetilde{P}_f(i+k-1, j+l-1) w_{kl},\ 1 \le i \le M,\ 1 \le j \le N, \tag{3}$$

where w_{kl} is the element of convolution kernel matrix W in the ith row and the jth column, and W is defined by

$$W = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

(3) L-layer

The L-layer consists of on–off nodes arranged in a $M \times N$ matrix form. All the nodes transform the light intensities of their counterparts in the R-layer into on–off signals which constitute matrix $Y(f)$, relying upon a shunting inhibition mechanism (Levine, 2018), namely

$$\frac{dY_{ij}(t)}{dt} = -\alpha Y_{ij}(t) + (\beta - Y_{ij}(t)) \times (R_{ij}(t) - g(R_{ij}(t))), \quad 1 \le i \le M, 1 \le j \le N, \tag{4}$$

with positive decay factor α, where $g(.)$ is a time-delay function, and β the maximal amplitude of excitation. Eq. (4) indicates that, once $Y_{ij}(t)$ increases with time t, the node of (i, j) will become excitatory, and conversely cause inhibition, which characterizes the visual response properties of the on–off nodes.

(4) M-layer

The M-layer only detects the change of visual motion direction in the left region (Region 3) in Fig. 3. Once some object(s) moves toward the left-down direction, LDN will become excitatory, and later generate a high membrane potential. Conversely, it will remain no excitation. More precisely, the M-layer consists of $M \times N$ medulla nodes distributed in a matrix form. Here, only those nodes in Region 3 are used to detect the change of local motion information. Accordingly, the direction potential matrix M_f, generated by the nodes in the M-layer is defined by

$$M_f(i,j) = \begin{cases} \sum_{k,l=0}^{2} abs(Y_{(i+k-1)(j+l-1)}(f)), & if \ \ 1 \le j \le n_L(i), \\ 0, & else, \end{cases} \tag{5}$$

with $1 \le i \le M$, where $n_L(i)$ is the number of medulla nodes in the ith row included in Region 3.

(5) LDN

LDN only responds to those medulla nodes in Region 3 in Fig. 3, and outputs a visual synthesized light intensity to characterize visual motion-directional information. Namely, it first gathers the activities of the nodes in Region 3 by

$$Sum = w \sum_{i=1}^{M} \sum_{j=1}^{n_L(i)} M_f(i,j), \tag{6}$$

where w is an adjustable positive parameter. Second, the output of LDN, which reflects the intensity of its response to those visual stimuli, can be acquired by

$$N_f^L = 2\left(\left(1 + \exp\left(-\frac{Sum}{m_L}\right)\right)^{-1} - 0.5\right), \tag{7}$$

where m_L represents the number of medulla nodes in Region 3.

Additionally, D-FVNN and R-FVNN also all include four neural layers and one output neuron. Their former three layers (i.e. P-, R- and L-layers) are completely the same as those in L-FVNN. The difference between L-FVNN and either D-FVNN or R-FVNN is given below:

(a) In D-FVNN, the M-layer, designed in a similar manner to that in L-FVNN highlights the change of visual motion in the front region which matches with Region 2 in Fig. 3. $n_L(i)$ and m_L in Eqs. (5)–(7) are here replaced respectively by the column number $n_D(i)$ at the last node in the ith row and the total of nodes in the region (i.e. m_D); the superscript of j takes values within $n_L(i) + 1$ and $n_D(i)$; N_f^L is displaced by N_f^D.

(b) The M-layer in R-FVNN only focuses on the right region which corresponds to Region 1 in Fig. 3. Herein, $n_L(i)$, m_L and N_f^L in Eqs. (5)–(7) are updated by $n_R(i)$, m_R and N_f^R, respectively; the superscript of j ranges from $n_R(i) + 1$ to N, and m_R is the total of nodes in Region 1.

4.3. Postsynaptic neural network

The postsynaptic neural network as in Fig. 2(b) is composed of four neurons, i.e. LDN, DDN, RDN and SDN as mentioned in Section 3.2. As the motion-directional detection neurons in the presynaptic neural structure of the fly, the former three neurons not only generate their spike-like potentials in terms of the external visual stimuli, but also send such potentials to the fly's postsynaptic neural network. Herein, the latter neuron as a spike detection neuron, i.e. SDN, is used to collect the outputs of the former three neurons, while transmitting alarm signals if possible danger happens. Once one or more moving objects appear in the wide field of view, at least one of the former three neurons will generate internal spikes by a common spike frequency mechanism. Specifically, each of the former three neurons executes the 0–1 spike response mechanism below,

$$S_f^i = \begin{cases} 1, & if \ \ \min(N_f^L, N_f^D, N_f^R) < T_s < \max(N_f^L, N_f^D, N_f^R), \\ 0, & else, \end{cases} \tag{8}$$

with $i \in \{L, D, R\}$, where T_s is the medium one of N_f^L, N_f^D and N_f^R as in Section 4.2. Afterwards, the activity of each neuron is determined by

$$\widetilde{N}_f^i = \begin{cases} N_f^i, & if \ \ T_s \le N_f^i \wedge \sum_{j=0}^{n_c} S_{f-j}^i(k) \le n_s, \\ 0, & else, \end{cases} \tag{9}$$

with spike threshold n_s. Accordingly, SDN gathers the outputs of the three neurons and later, creates an excitatory intensity at frame f by

$$V_f^{LP} = \begin{cases} \widetilde{N}_f^L, & if \quad \widetilde{N}_f^L > \max(\widetilde{N}_f^D, \widetilde{N}_f^R), \\ \widetilde{N}_f^D, & if \quad \widetilde{N}_f^D > \max(\widetilde{N}_f^L, \widetilde{N}_f^R), \\ \widetilde{N}_f^R, & if \quad \widetilde{N}_f^R > \max(\widetilde{N}_f^L, \widetilde{N}_f^D), \\ 0, & else. \end{cases} \tag{10}$$

5. Fly visual synthesized collision detection model (FVSCDM)

This section first designs FVSCDM to perform multiple-regional collision detection in Section 5.1. Second, FVSCDM's algorithmic formulation and the related computational complexity are given in Sect.5.2.

5.1. Model design on FVSCDM

In terms of Fig. 4, FVSCDM consists of AFVJPNN as in Section 4 and three detection mechanisms of feedforward inhibition(FFI), synthesized spike inhibition(SSI) and collision detection (CD). Herein, AFVJPNN is used to receive successive visual images and generates the related membrane potentials of neuron SDN which present the change of the visual scene. The FFI mechanism, based on the previous works (Mauss, Meier, Serbe, et al., 2014; Wang, Dewell et al., 2018; Wang, Peng et al., 2018) is designed to prevent from false collision warning when some moving object(s) suddenly oscillates in the visual scene. The SSI mechanism integrates SDN with FFI, and generates a possible spike signal. After that, the CD mechanism outputs a possible alarm by multiple adjacent visual frames.

(1) FFI

FFI first collects the average of grayscales at the fth frame, while generating a smoothed gray value by

$$F(f) = V^{FFI}(f) + \sum_{l=1}^{n_c} \kappa_l V^{FFI}(f - l), \tag{11}$$

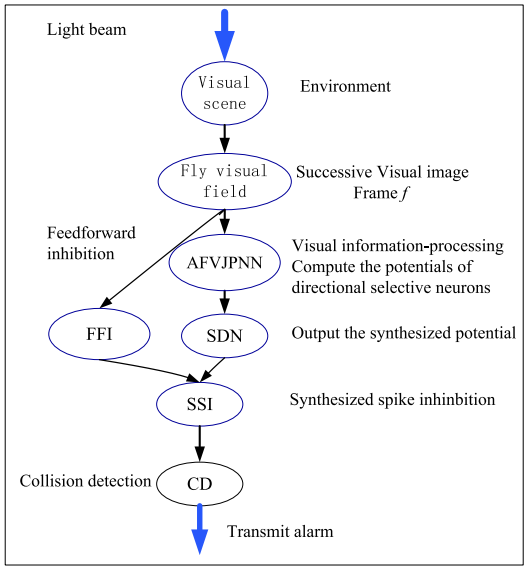


Fig. 4. Schematic framework on FVSCDM.

where

$$V^{FFI}(f) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} V_f(i,j). \tag{12}$$

Second, it utilizes a first-order low-pass filter (Fu, Bellotto et al., 2018; Fu, Hu et al., 2018; Fu et al., 2019) to eliminate noise information by

$$\frac{d\widetilde{F}(t)}{dt} = \frac{1}{\tau}(F(t) - \widetilde{F}(t)), \tag{13}$$

with positive speed attenuation factor τ, where $\widetilde{F}(t)$ represents the output of FFI at the moment of t. Finally, a spike inhibition scheme is designed to detect whether FFI delivers a spiking signal at frame f by

$$S_f^{spike} = \begin{cases} 1, & if\ \widetilde{F}(f) \le T_{PF}, \\ 0, & else, \end{cases} \tag{14}$$

where T_{PF} is a dynamic threshold (Yue & Rind, 2006a, 2006b), i.e. $T_{PF} = T_0 + k_F \times \widetilde{F}(f)$ with positive T_0 and k_F.

(2) SSI

After receiving the output of FFI and the activity of SDN at the moment f, SSI as a spike detector produces a synthesized collision warning signal to display the status of the visual scene by

$$CDN_f = \begin{cases} \widetilde{F}(f), & if\ \sum_{i=0}^{n_c} S_{f-i}^{spike} \ge n_s, \\ V_f^{LP}, & else, \end{cases} \tag{15}$$

with spike threshold n_s, by which a spike inhibition scheme can be developed by

$$S_f^{CDN} = \begin{cases} 1, & if\ CDN_f \ge T_f, \\ 0, & else, \end{cases} \tag{16}$$

where T_f is determined by one dynamical threshold scheme (Zhang et al., 2015).

(3) CD

CD as an alarm-transmitting detector is designed to ultimately decide whether a collision alarm is transmitted by means of SSI at frame f. Specifically, once SSI derives n_s successive spikes, a possible danger will occur, and hence delivers a collision warning signal by

$$Alarm_f = \begin{cases} 1, & if\ \sum_{l=0}^{n_s} S_{f-l}^{CDN} \ge n_s, \\ 0, & else. \end{cases} \tag{17}$$

5.2. Algorithm formulation and computational complexity

As related to the structure of AFVJPNN in Fig. 2(b) and the above designs of Eqs. (1)–(17), FVSCDM's algorithm is formulated below.

Algorithm 1 Multi-regional collision detection on FVSCDM
Input: Grayscale frame V_f;
Output: N_f^L, N_f^D, N_f^R, T_f and $Alarm_f$.
Step 1: Parameters setting of τ, n_s, n_c and w;
Step 2: Compute grayscale matrix P_f at frame f by Eq. (2) after the ViBe method is enforced on V_f;
Step 3: Generate L-FVNN's membrane potential and output N_f^L by Eqs. (1)–(7); likewise, compute N_f^D and N_f^R for D-FVNN and R-FVNN, respectively;
Step 4: Compute SDN's excitatory intensity V_f^{LP} by Eqs. (8)–(10);
Step 5: Utilize Eqs. (11)–(14) to judge whether a spike occurs at frame f inside FFI;
Step 6: Decide whether an early warning signal arises at frame f by Eqs. (15)–(17);
Step 7: Stop the process of collision detection only when no requirement is satisfied, and return to step 2 otherwise.

FVSCDM's computational complexity is mainly decided by steps 2–6. Within a period of execution, the ViBe method extracts the foreground of the current grayscale image of V_f, which needs to execute M_1 times with $M_1 = 2MN$. Furthermore, since each sub-network needs to execute at most M_2 arithmetic operations with $M_2 = (46 + 2n_c)MN + 6$, step 3 runs at most $3M_2$ times. Step 4 runs M_3 operations with $M_3 = (n_c + 3)MN + 8$. Step 5 needs M_4 operations to judge whether a spike takes place with $M_4 = 2(MN + n_c) + 7$. Additionally, in order that FVSCDM can send early warning signals, step 6 needs to compute the dynamic threshold of T_f in Eq. (17) with at most n_c^2 times, and also executes M_5 arithmetic operations with $M_5 = 3n_c + 2$. Summarily, steps 2–6 run at most M_6 times in a period of time, where

$$M_6 = (7n_c + 145)MN + 6n_c + n_c^2 + 35.$$

Since n_c is defined as a small time step, FVSCDM's complexity in the worst case can be determined by $O(145MN)$.

6. Experimental study

To examine whether AFVJPNN-based FVSCDM can effectively execute multiple-regional collision detection, the environmental setting is given in Section 6.1. Section 6.2 detects whether AFVJPNN as in Section 4 can present preferential response properties based on eight kinds of video sequences generated by a football. AFVJPNN's performance test is performed in Section 6.3 by virtue of three kinds of video sequences from two visual scenes. Three reported collision detection models mentioned in Fig. 1 are picked up to execute comparative analysis on performance effectiveness in Section 6.4, relying upon four video sequences. Finally, FVSCDM's efficiency is analyzed in Section 6.5.

Table 1
Parameter settings of AFVJPNN and FVSCDM.

Name	Value	Name	Value
Excitatory threshold T_c	0.7	Decay factor α	5
Maximal amplitude β	10	Spike threshold n_s	5
Adjustable parameter w	0.1∼0.9	Speed attenuation τ	0.02
Time step n_c	5		

6.1. Environmental setting

Throughout the whole experimental study, all experiments are executed on a Windows 10 system with i7CPU/3.60 GHz and RAM/16G. Source codes are written in Visual Studio 2013 with Open CV3.4.0. The sample videos below are shot by us based on Sony HDR-CX700 or generated by computer. Each video sequence is recorded at a frame rate of 50fps, and later separated into 8-bit grayscale images. After parameters tuning, we define FVSCDM's parameters given in Table 1, based on the previous works (Fu, Hu et al., 2018; Missler & Kamangar, 1995; Yue & Rind, 2006a, 2006b; Zhang et al., 2015).

6.2. Performance test on L-, D- and R-FVNNs

To sufficiently examine whether AFVJPNN can effectively adapt to different kinds of visual stimuli, eight video sequences, acquired by computer-generated visual stimulation are hereby taken to analyze whether they can respond to preferential visual regions.

Case I: Radial motion

Four series of sample frames in Fig. 5 below are chosen to detect whether the three subnetworks of AFVJPNN can preferentially respond to motion-directional stimuli. Each video sequence includes 500 frames with size 110 × 60 per frame. Fig. 5(a)–(c) indicates that the same football generates three different scenes, namely the football slowly moves from the top-left, up-front or top-right corner to the camera, and hence its contour gradually expands. Fig. 5(a), related to the visual region division as same as that in Fig. 3 indicates that the football looms in Regions 2 and 3; in Fig. 5(b), the football moves toward the camera only in Region 2, while in Fig. 5(c), the same is true in Regions 1 and 2. Besides, Fig. 5(d) shows that three footballs of the same size simultaneously approach the camera along different directions.

Figs. 5(a) and 6(a) show that, when the football approaches the camera in Region 3 from the top-left corner to the left-bottom (i.e. frames 50–300), only L-FVNN is activated after 50 frames, and gradually generates high membrane potentials. Once it looms in Regions 2 and 3 (i.e. frames 300–500), L- and D-FVNNs all become excitatory. R-FVNN, however, has no response to one such visual scene, since the football does not appear in Region 1. Similarly, by Figs. 5(b) and 6(b) only D-FVNN is triggered to yield high membrane potentials, since the football moves almost only in Region 2 and approaches the camera gradually. Likewise, Figs. 4(c) and 5(c) validate that, when the football scrolls in Region 1, only R-FVNN makes a response, and also gains high excitatory intensities if it is close to the camera (i.e. frames 50–340). Particularly, once the football scrolls in Regions 1 and 2, D- and R-FVNNs can result in excitation. All these show that the three sub-networks can selectively respond to their preferred regions while L- and R-FVNNs can execute edge detection. This is assistant with the functional properties of their output neurons, namely such neurons are particularly responsive to the left, front and right regions in the field of view, respectively.

Additionally, Figs. 5(d) and 6(d) illustrate that, once multiple footballs synchronously move toward the camera from different directions, the sub-networks can be simultaneously activated after 50 frames, while yielding high excitatory intensities increasingly. In such a case, their output membrane potential curves almost synchronously change within frames 51 and 470. Afterwards, since the left and right footballs have no contour change, L- and R-FVNNs degrade. D-FVNN, however, keeps excitatory, as the middle football is still active. These validate that the above three sub-networks can synergistically detect all changes in the wide field of view.

Case II: Curve movement

Four visual sequences with size 110 × 60 per frame in Fig. 7 are chosen to check whether the above sub-networks can also preferentially respond to non-radial motion patterns. The visual sequences have different frame numbers (i.e. 1050, 1050, 900 and 1500), due to the requirement of different types of motion patterns. Fig. 7(a) shows that a football makes periodical movement on the line segment. It directly moves in Region 2 from point B to point A within frames 51 and 250, and later immediately recedes from point A to point B within frames 251 and 554. This process is repeated twice. Fig. 7(b) shows that a football periodically makes counterclockwise rectangular movement after starting from the up-front point, which indicates that it walks in Regions 1, 2 and 3. Fig. 7(c) shows that a football starts at the up-front point near the ground, and then rolls along the S-type curve. Thereby, it passes through Regions 1, 2 and 3. In Fig. 7(d), a football starts at the up-front point, and then moves freely. It rolls orderly through points A (frame 925), B, C, D and E (frame 1450), and thus moves through each of the three regions with many times. These visual sequences can be adopted to sufficiently detect whether the sub-networks can adapt to complex visual scenes.

Figs. 7(a1) and 8(a) validate that, once the football moves radially from point B to point A in the front of the field of view, only D-FVNN becomes excitatory and generates high membrane potentials increasingly. When the football recedes from point A to point B, D-FVNN degrades, and correspondingly its membrane potential decreases with time. These reveal that D-FVNN has the property of directional selectivity. In one such case, L- and R-FVNNs keep no response, since there is no local motion information in Regions 1 and 3. Figs. 7(b1) and 8(b) illustrate that, since the football initially scrolls along the left-up line segment, L-FVNN remains excitatory within frames 51 and 180. Afterwards, D-FVNN becomes excitatory within frames 450 and 700 or within frames 980 and 1050. R-FVNN only keeps excitatory within frames 550 and 980. This indicates that the sub-networks can get with the football's motion change. Figs. 7(c1) and 8(c) illustrate that, when the football moves along the S-type curve in the field of view, the sub-networks are all activated, since Regions 1, 2 and 3 include its motion information. Particularly, they do not make any excitatory response to their non-preference regions. This illustrates that they can synergistically detect all possible changes in the whole field of view.

Figs. 7(d1) and 8(d) show that, when the football moves freely, the sub-networks can only generate irregular membrane potential curves. However, they also present some clear features, for example they hardly respond to the football within frames 1 and 710, since the football is far from the camera. When the football appears in Regions 2 and 3 (frames 750–850), L- and D-FVNNs are activated and generate high membrane potentials. Hereafter, since it almost makes slow shift within frames 1100 and 1230, the sub-networks almost keep no response. However, once the football scrolls in Regions 1 and 2 at a fast speed, D- and R-FVNNs can persistently keep either excitation or ipsilateral horizontal

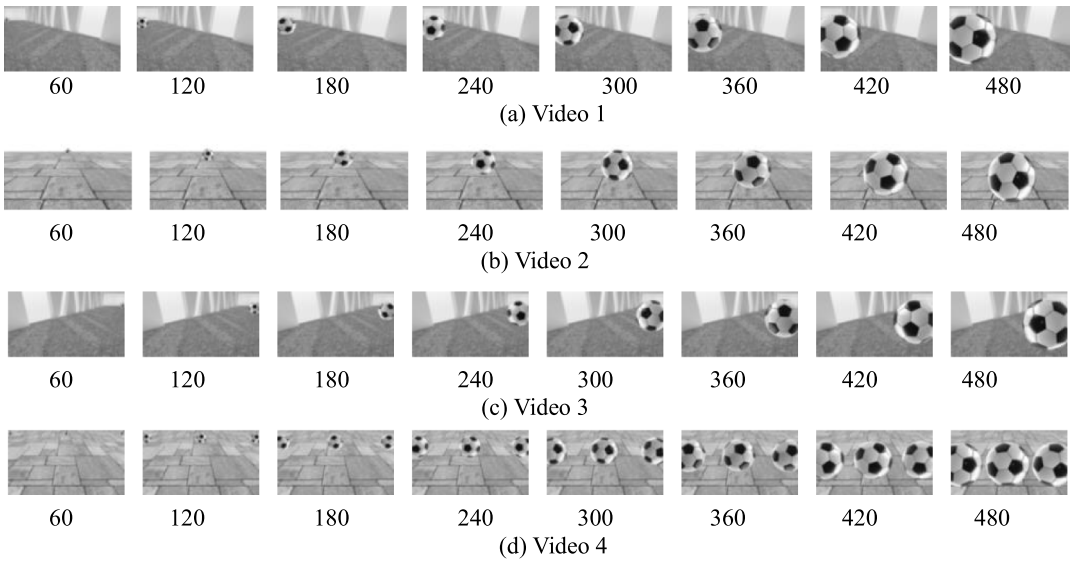


Fig. 5. Illustrative example frames. Each video sequence is illustrated only by eight frames; the frame number is indicated under each image. (a) A football moves from the top-left corner to the camera; (b) a football scrolls from the up-front corner to the camera; (c) a football moves from the top-right corner to the camera; (d) three footballs approach the camera simultaneously at the same speed along different directions (top-left, up-front, and top-right).

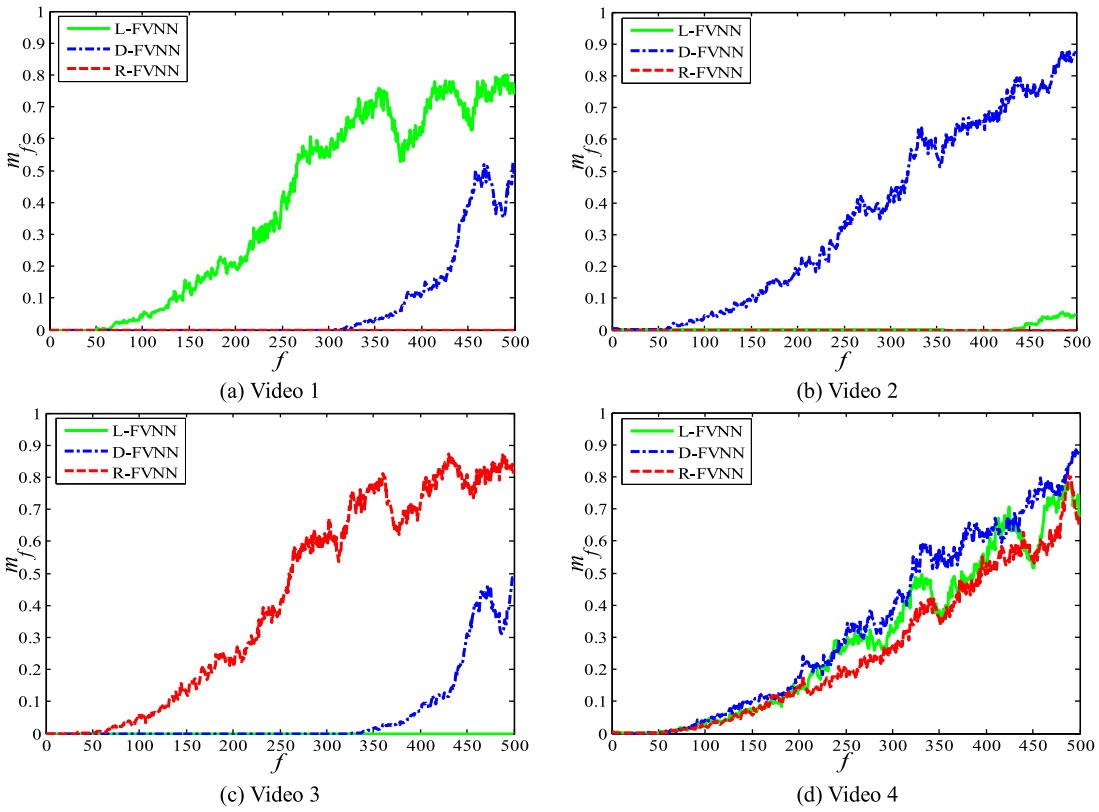


Fig. 6. FVNN's response curves related to the matched visual scenes in Fig. 5, where m_f stands for the membrane potential of a sub-network at frame f.

inhibition. These show that the sub-networks can adapt to the change of directional motion of the football, and especially L- and R-FVNNs can perform edge detection.

6.3. Performance test on AFVJPNN

To check whether AFVJPNN can effectively adapt to different kinds of actual scenes, we first analyze how different image-processing approaches influence AFVJPNN's performance. Second, its performance characteristics are also discussed.

Case I: Image preprocessing

The inter-frame difference (IFD) approach (Weng et al., 2010) as an object extraction approach is here picked up to preprocess FVNN's input frames, in order to justify that the ViBe approach (Barnich & Van Droogenbroeck, 2011) is more effective in preventing AFVJPNN's collision detection performance from being influenced by background noise. Meanwhile, the visual scene with 200 frames and each frame size of 110 × 60 in Fig. 9(a) is here selected to detect how IFD and ViBe affect the response

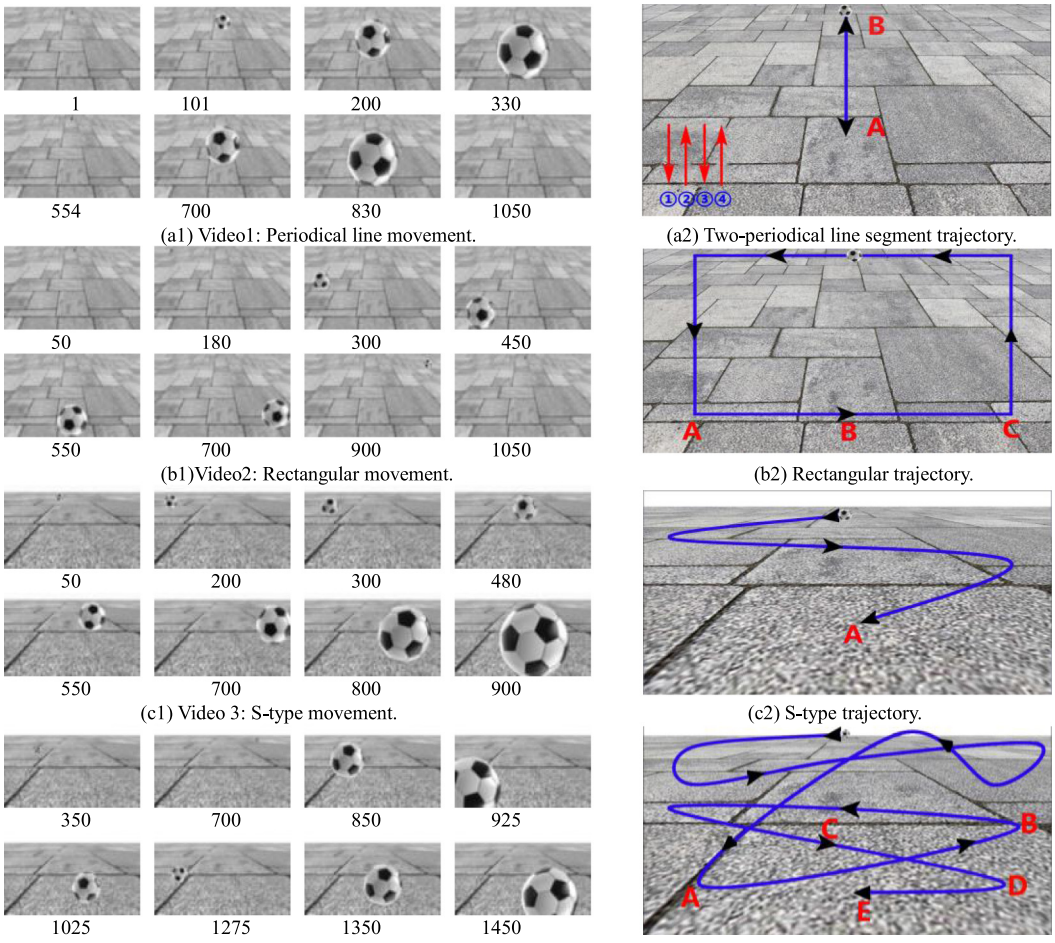


Fig. 7. Four video frame sequences of the football and the correspondingly related movement trajectories. Each video sequence is illustrated only by eight frames; the frame number is indicated under each image.

performances of the three sub-networks and AFVJPNN. In the scene, a black bucket starts from the far end in the front of the field of view, and subsequently scrolls to the bottom along the left-down direction after 195 frames. Therefore, it necessarily passes through Regions 2 and 3. Fig. 9(b) and (c) can clearly validate that ViBe outperforms IFD, as the former one can more perfectly execute object extraction.

As can be seen from Fig. 10, IFD and ViBe can make the sub-networks and AFVJPNN present different response characteristics. With the help of ViBe, L-FVNN and AFVJPNN produce extremely high membrane potentials when the bucket approaches the left-bottom gradually, and meanwhile can also be activated in the case where the bucket moves in Region 3. However, IFD makes L-FVNN and AFVJPNN mistakenly acquire high membrane potentials within frames 185 and 190, due to noisy interference. Thereby, it is necessary to preprocess input frames for the sub-networks and AFVJPNN, while ViBe is an alternative tool for the current work.

Case II: Excitatory intensity change

Two visual scenes given in Fig. 11, captured by our camera are taken to detect whether AFVJPNN's response property is influenced by different kinds of moving objects. With a total of 1000 frames and a size of 110 × 60 per frame, the first sequence in Fig. 11(a) describes the process of which a pedestrian walks at a constant speed and ultimately appears in the front of the camera. The second video sequence in Fig. 11(b) consists of 650 frames with size 110 × 60 per frame, in which a car as a large object approaches the camera along the wide road at a fast speed.

Fig. 11(c) can distinctively illustrates that, when the pedestrian radially walks in the distal top-upper direction (frames 1–700), AFVJPNN can only generate small membrane potentials. However, once the pedestrian appears in the front of the visual region after 700 frames, AFVJPNN will increasingly acquire high membrane potentials by D-FVNN. Moreover, when the car drives within frames 350 and 700, AFVJPNN can only acquire low membrane potentials, and later it yields high excitatory values by D- and R-FVNNs within frames 700 and 900. Summarily, AFVJPNN can adapt to different visual scenes by automatically switching between the three sub-networks.

6.4. Collision detection and comparative analysis

To evaluate FVSCDM's collision detection performance and its success rate on early collision warning, four visual scenarios with resolution 110 × 60 in Fig. 12 are here taken into account. First, we detect whether FVSCDM can discover the collision regions included in the video sequences in terms of one dynamical threshold scheme (i.e. T_f) mentioned in Eq. (16). Second, three typical visual collision detection models given in Fig. 1 are chosen to participate in comparison with FVSCDM. AFVNN receives images with resolution 128 × 128 because of its inherent structure, but so do LGMD and LGMD1 with image resolution of 110 × 60.

Case I: Comparative analysis on collision region

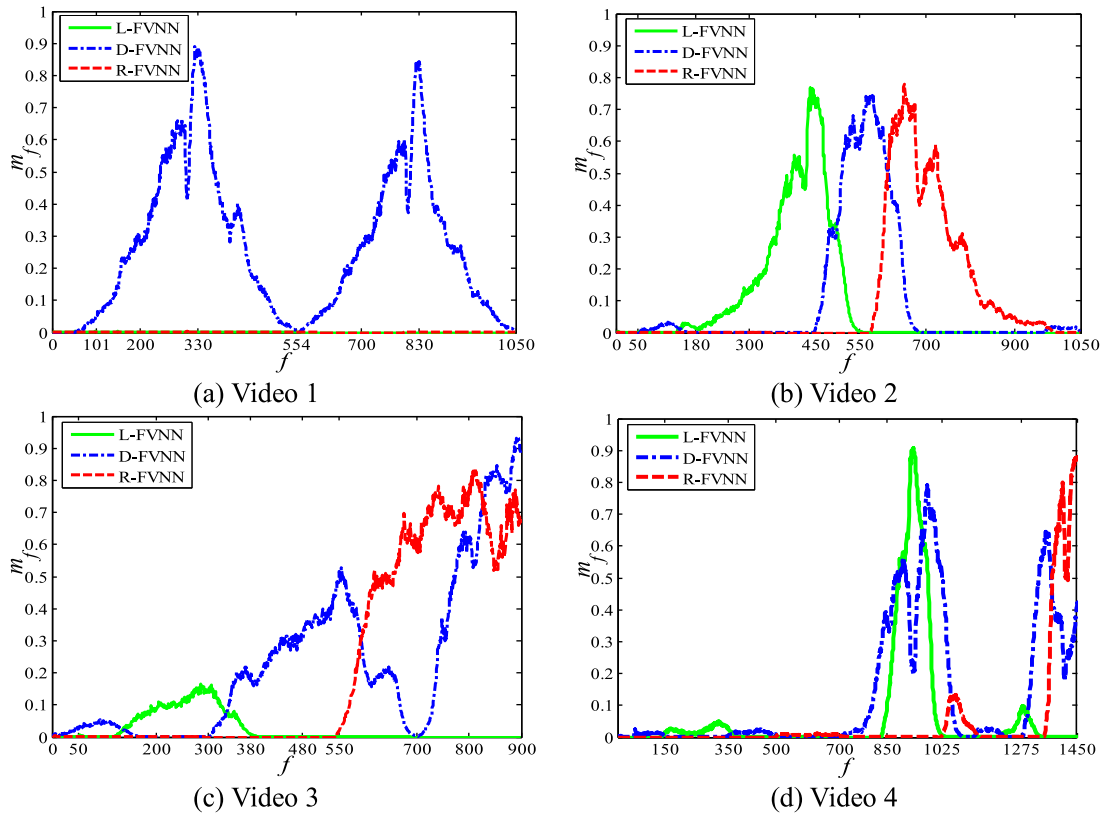


Fig. 8. FVNN's response curves related to the corresponding video sequences in Fig. 7.

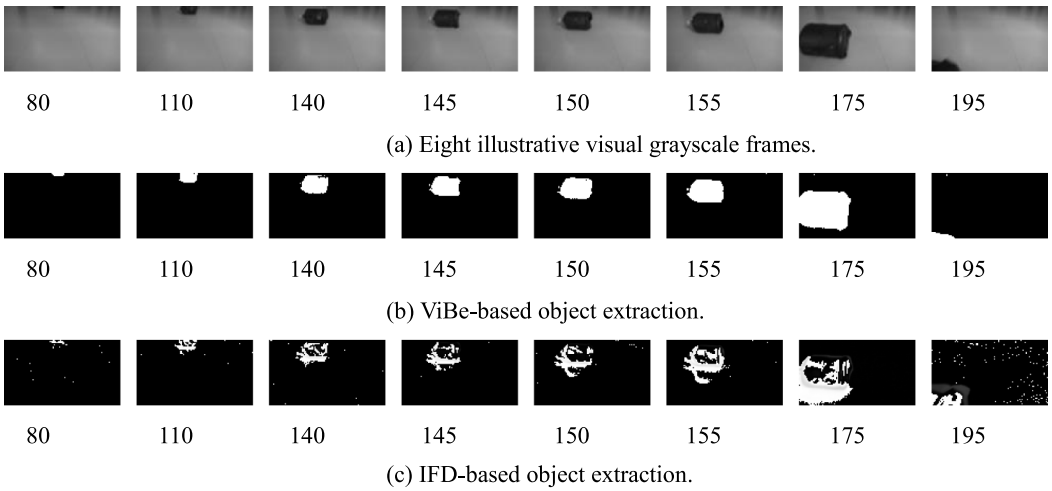


Fig. 9. Illustrative frames and extraction quality.

Fig. 12(a) illustrates that a basketball directly scrolls on the ground toward the camera at a constant speed. As related to the division of the field of view as in Fig. 3, the basketball starts from the up-front point, and scrolls among Regions 1, 2 and 3. In such a case, collision signals will occur within frames 80 and 90. Fig. 12(b) describes that the car directly drives toward the camera at a fast speed, and certainly its contour expands rapidly so that such three regions are almost sheltered by the car. Therein, the collision region is included within frames 547 and 557. Additionally, Fig. 12(c) shows that, after the autotruck initially appears at the distal top-left corner, it drives toward the right-down corner along the right diagonal line at a fast speed. In one such case, the autotruck moves in the three regions, while early warning signals take place within frames 142 and 220. Besides, Fig. 12(d) presents a motion scene where two cars move along the front-to-after direction. Herein, the first car moves along the left-down direction after directly driving 80 frames toward the camera, and thus possible collision happens within frames 95 and 105. Subsequently, the second car appears in the top-front field of view, but does not result in collision. Thereby, the two cars drive in Regions 2 and 3.

Figs. 12(a) and 13(a) validate that, when the basketball as a small moving object is about to reach the camera, FVSCDM produces a high membrane potential curve segment between frames 80 and 90 (i.e. collision region) that exceeds the threshold curve. This shows that FVSCDM can discover the basketball's collision region. It is pointed out that the membrane potential curve begins to descend slightly when the basketball moves from frame 81

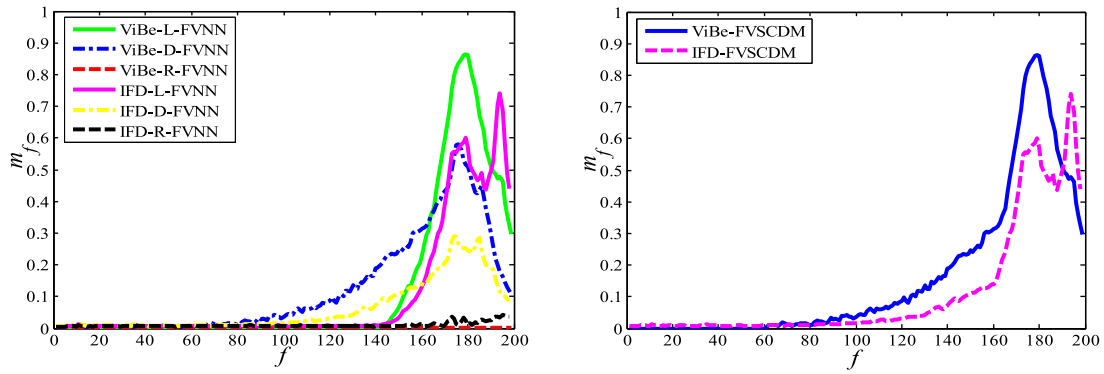


Fig. 10. Comparison of response curves related to IFD and ViBe. (a) Response curves of the sub-networks, in which X-Y-FVNN represents X-based sub-network Y-FVNN with X ∈ {IFD,ViBe} and Y ∈ {L, D, R}; (b) response curves of AFVJPNN; X- AFVJPNN stands for X-based AFVJPNN.

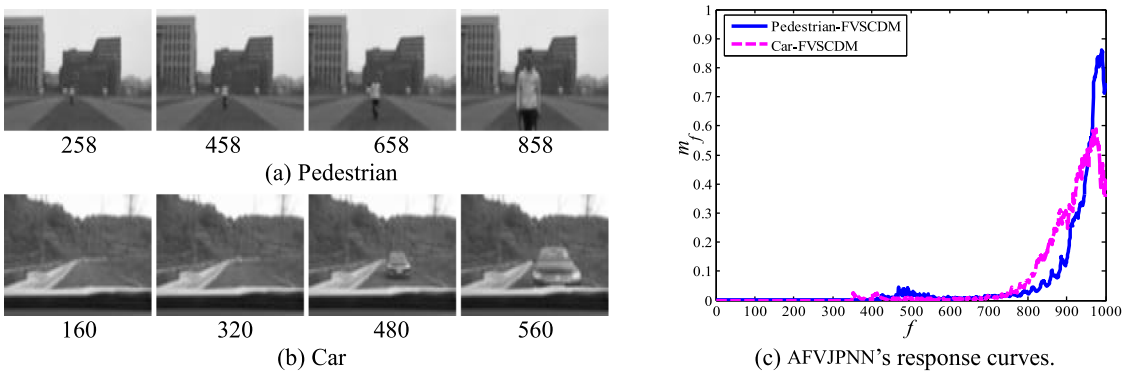


Fig. 11. Illustrative example frames and response curves; the frame number is indicated under each image. (a) Pedestrian's radial movement; (b) car's radial movement; (c) AFVJPNN's response curves related to the two video sequences.

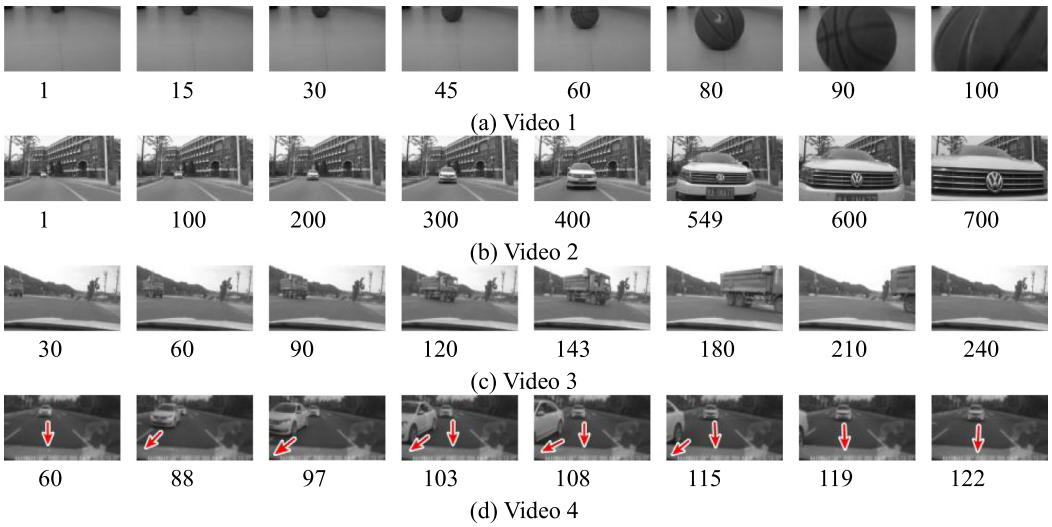
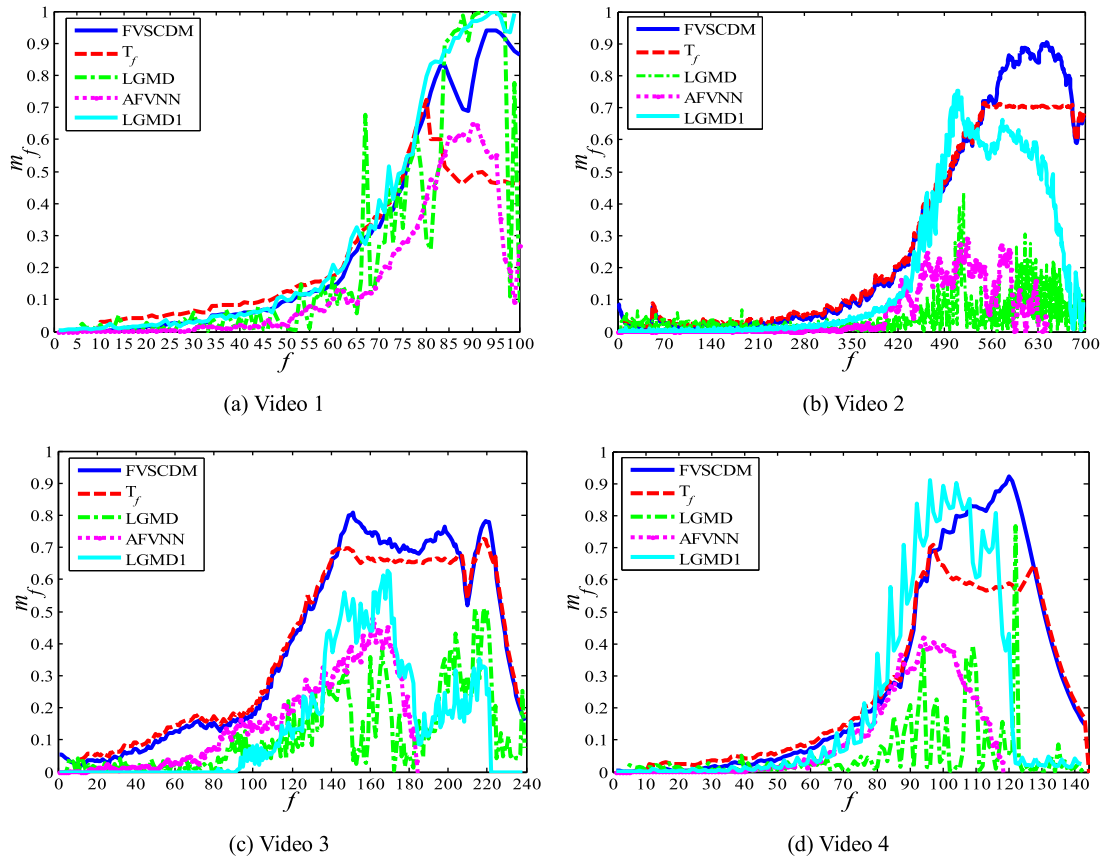


Fig. 12. Illustrative example frames of four video sequences, each of which is illustrated by eight frames. (a) Basketball's radial movement; (b) car's radial movement; (c) autotruck's diagonal movement; (d) two cars' radial movement.

to frame 86. The main reason consists in that, when it deviates from Region 2 and moves toward Region 1, FVNN switches from D-FVNN to R-FVNN. Additionally, AFVNN can generate high membrane potentials within frames 85 and 95, and thus a narrow collision region can be found; LGMD and LGMD1 can also find their wide collision regions [83,93] and [79,89], respectively, for which the main reason is because their membrane potential curves are located beyond the threshold curve segment. Similarly, as related to Figs. 12(b), 13(b) illustrates that FVSCDM can accurately find the collision region (i.e. [547,557]), and its membrane potential curve almost keeps stable change. The compared models, however, cannot discover the collision region except for LGMD1. It is highlighted that, since the background in Fig. 12(b) causes strong noise signals, the inter-difference method, i.e. IFD,



(a) Video 1 (b) Video 2

(c) Video 3 (d) Video 4

Fig. 13. Collision region and the comparison of membrane potential curves acquired by the four models related to Fig. 12; T_f represents the dynamic value at frame f mentioned in Eq. Eq. (16); the collision region is denoted by a closed interval in which the values of T_f are beyond 0.7.

can only make LGMD1 produce relatively high membrane potentials and an extremely narrow collision region near frame 505, due to strong noise signals generated by the background in Fig. 12(b).

Fig. 13(c) shows that, when the autotruck moves diagonally from the distal top-left corner to the right-down corner at a fast speed, the above four models exhibit distinctive response features. The compared models do not make strong reaction, and hence cannot find any collision region. When the autotruck appears in the distal front of the field of view, LGMD1 cannot trigger early collision warning, even if it can get relatively high membrane potentials from frame 135 to frame 170. FVSCDM, however, can persistently produce high membrane potentials from frame 142 to frame 220 in terms of its R-FVNN, and thereby a wide collision region can be detected by virtue of the threshold curve. This indicates that it can rationally detect edge information. As associated to Figs. 12(d), 13(d) hints that, when the first car approaches the left-bottom edge from Region 2 to Region 3, L- and D-FVNNs as in FVSCDM are strongly activated with high membrane potentials. Afterwards, although such a car vanishes from such two sub-regions, D-FVNN can make weak response, since the second car appears in the far front of the camera. Thereby, FVSCDM can continually yield high membrane potentials within frame 95 and 120 by means of L- and D-FVNNs, and hence can capture a collision region, i.e. [95,120]. LGMD1 can find a collision region [90,100], but cannot well respond to the second car in the distal top-front field of view. However, since AFVNN and LGMD can only output low membrane potentials, they cannot discover any collision region.

Case II: Comparative analysis on collision detection's success rate

In the table below, the fundamental materials are given in the left table, and meanwhile the experimentally statistical results are displayed in the right table, depending on the experiment in the above case.

Table 2 presents some statistical results on early collision warning and trigger frames for the four video sequences in Fig. 12, where we define a frame as a triggering frame if the area of the frame is three times that of the object. We can clearly draw three conclusions:

A. LGMD and AFVNN can only get a low success rate for the above four scenes, as their membrane potential curves are almost below the dynamic threshold curves. LGMD1 can acquire a relatively high success rate, but background noises easily influence its collision detection performance. FVSCDM can successfully transmit collision alarms for the above four scenes.
B. FVSCDM can well perform multiple-regional collision detection and edge detection when one or more moving objects appear in the whole field of view, as it executes collision detection by switching L-, D- and R-FVNNs. However, the compared models are difficult.
C. FVSCDM can capture the trigger frames in the four video sequences in Fig. 12, whereas the compared models and in particular LGMD1 cause early or delayed collision warning.

Summarily, when one or more objects move in the wide field of view, FVSCDM can well carry out collision detection and in particular multiple-regional collision detection with a high success rate, while its L- and R-FVNNs can rationally detect the change of luminance intensity in the left and right regions. LGMD1 is a competitive collision detection model, even if its membrane potential curve in Fig. 13(b) and (c) is below the threshold curve. Especially, it can respond to a small object in the far front of

Table 2
Experimental statistics. Cars I and II stand for the video sequences in Fig. 12 (b) and (d), respectively; SR stands for the success rate that a collision detection model can transmit early collision warning within all the collision regions.

Actual scene information				Experimental result (collision regions and *SR*)					
Video type	Total of frames	Collision Yes/No	Trigger frame	Model	Basketball	Car I	Autotruck	Car II	SR (%)
Basketball	101	Yes	80	LGMD	83–93	N/A	N/A	N/A	25
Car I	700	Yes	549	LGMD1	79–89	505–515	N/A	90–100	50
Autotruck	240	Yes	143	AFVNN	85–95	N/A	N/A	N/A	25
Car II	144	Yes	97	FVSCDM	80–90	547–557	142–220	95–105	100

Table 3
Runtime statistics. *AT*(a) and *AT*(b) represent the average runtime of FVSCDM per frame with image resolutions 161 × 91 and 110 × 60, respectively.

Virtual scenes in Section 6.2				Practical scenes in Sections 6.3 and 6.4			
Video type	Total of frames	AT(a)	AT(b)	Video type	Total of frames	AT(a)	AT(b)
Fig. 5(a)	500	0.64	0.23	Fig. 11(a)	1000	0.64	0.23
Fig. 5(b)	500	0.66	0.24	Fig. 11(b)	650	0.67	0.24
Fig. 5(c)	500	0.67	0.24	Fig. 12(a)	101	0.68	0.25
Fig. 5(d)	500	0.65	0.24	Fig. 12(b)	700	0.62	0.23
Fig. 7(a1)	1050	0.65	0.24	Fig. 12(c)	240	0.68	0.25
Fig. 7(b1)	1050	0.64	0.24	Fig. 12(d)	144	0.66	0.26
Fig. 7(c1)	900	0.66	0.24	FVSCDM's average runtime per frame		0.65 s	0.24 s
Fig. 7(d1)	1500	0.64	0.23				

the field of view, but cannot effectively detect the change of visual motion in the left or right regions. Additionally, whereas AFVNN and LGMD can acquire their collision regions for video 1 in Fig. 12(a), they cannot correctly respond to the other visual scenes in Fig. 12.

6.5. FVSCDM's efficiency analysis

As related to Section 4.3, FVSCDM needs to compute the membrane potentials of the three sub-networks and decides whether a collision alarm transmits. Its efficiency is *decided* by steps 3–6 in Algorithm 1 in the process of information-processing for a single image frame. Hereby, we only evaluate the whole runtime spent by such four steps for a frame by means of the average runtime criterion below,

$$AT = \frac{1}{m}\sum_{f=1}^{m} t_f, \tag{18}$$

with image sample size m, where t_f stands for the four steps' runtime for the fth frame. AT is here regarded as FVSCDM's average runtime. We take the virtual and actual video sequences with resolutions of 161 × 91 and 110 × 60 in Figs. 4, 6, 10 and 11 as examples. The frame efficiency of FVSCDM is shown in Table 3.

Table 3 illustrates that, when FVSCDM processes virtual and actual videos, there does not exist clear efficiency difference, provided that the image frames are with the same resolution. More precisely, when each virtual or actual image frame is processed at a resolution of 161 × 91, steps 3–6 in FVSCDM take about 0.65 s, but only spend about 0.24 s if the resolution is 110 × 60. Thereby, FVSCDM's efficiency is not influenced by the type of visual scene, but extremely depends on images' resolution. Once a grayscale image is enlarged, FVSCDM will naturally spend much more runtime to execute such four steps. Summarily, in order that FVSCDM can acquire a higher success rate under a lower time cost, it is reasonable to take the image resolution of 110 × 60.

7. Conclusions and future works

With the rapid development of collision detection research, biologically-inspired visual information-processing mechanisms will be an irreplaceable bio-inspiration for developing artificial visual neural networks and collision detection models. Particularly, bio-visual synthesized collision detection will become an active topic in computer vision. Unfortunately, whereas single-regional collision detection as an old but active topic has attracted the attention of researchers, multiple-regional collision detection still remains a virgin territory. Therefore, an artificial fly visual joint perception neural network (AFVJPNN) capable of being applied to multiple-regional collision detection is proposed to describe the changes of visual motion in multiple visual sub-regions. Four prominent design inspirations can be summarized: (i) a parabolic curve is used to divide a visual region into three sub-regions in terms of visual perceptual characteristics; (ii) an image preprocessing approach is cited to ensure that AFVJPNN's performance is not influenced by background noises in input frames; (iii) AFVJPNN is designed to describe the performance characteristics of one or more moving objects in the wide field of view; and (iv) a multiple-regional collision detection model (FVSCDM), which includes AFVJPNN, is designed to online decide whether to transmit early collision warning immediately.

FVSCDM possesses some striking merits such as structural simplicity and multiple-regional synchronous collision detection, while L- and R-FVNNs in AFVJPNN can carry out edge detection. Experimentally, the performance test and comparative analyses can conclude four points: (i) the fly visual system as a biological organism is an invaluable bio-inspiration for engineering applications such as early collision warning and navigation; (ii) L-, D- and R-FVNNs in AFVJPNN themselves have distinctive preferences to specific motion-directional clues; (iii) FVSCDM can effectively measure the performance characteristics of one or more moving objects presented in the *whole* field of view, since AFVJPNN is achieved by the three sub-networks; and (iv) FVSCDM with a high success rate, which only spends about 0.24 s per frame for collision detection, can distinctively perform well over the compared visual neural networks.

Although the current work is devoted to exploring AFVJPNN and FVSCDM in order to online solve the problem of collision detection in the wide field of view, some issues need to be further studied. For example, in order to ensure that FVSCDM can effectively enforce collision detection, its intrinsic parameter settings need to satisfy different kinds of visual scenes automatically; the robustness of its object detection needs to be further enhanced. Whereas the present work initially touches upon multi-object collision detection, it is still open how to

develop omni-directional collision detection models based on fly neural systems. In the future work, we will investigate an automatic avoidance system, based on FVSCDM and bio-inspirations from the neurophysiology of the fly's brain, while transplanting it into a control processor in order to develop an anti-collision robot.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors sincerely thank the anonymous reviewers for their valuable comments. The authors also thank the editors of this work for their support. The work is supported by National Natural Science Foundation of China (No. 62063002, 61563009), and Science and Technology Planning Project of Guizhou Province, China ([2018] 5781–48).

References

Aptekar, J. W., Shoemaker, P. A., & Frye, M. A. (2012). Figure tracking by flies is supported by parallel visual streams. *Current Biology*, *22*(6), 482–487.
Barnich, O., & Van Droogenbroeck, M. (2011). ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, *20*(6), 1709–1724.
Bishop, L. G., & Keehn, D. G. (1966). Two types of neurons sensitive to motion in the optic lobe of the fly. *Nature*, *212*(5068), 1374.
Blanchard, M., Rind, F. C., Verschure, P. F., et al. (2000). Collision avoidance using a model of the locust LGMD neuron. *Robotics and Autonomous Systems*, *30*(1), 17–38.
Borst, A., Haag, J., Reiff, D. F., et al. (2010). Fly motion vision. *Annual Review of Neuroscience*, *33*(1), 49–70.
Buchner, E., Buchner, S., & Bülthoff, I. (1984). Deoxyglucose mapping of nervous activity induced in Drosophila brain by visual movement. *Journal of Comparative Physiology*, *155*(4), 471–483.
Clark, D. A., Bursztyn, L., Horowitz, M. A., et al. (2011). Defining the computational structure of the motion detector in Drosophila. *Neuron*, *70*(6), 1165–1177.
Clifford, C. W., & Ibbotson, M. R. (2002). Fundamental mechanisms of visual motion detection: models, cells and functions. *Progress in Neurobiology*, *68*(6), 409–437.
Cuntz, H. (2014). Fly lobula plate tangential cells (LPTCs), models of. In D. Jaeger, & R. Jung (Eds.), *Encyclopedia of computational neuroscience*. New York, NY: Springer.
Dalgaty, T., Vianello, E., De Salvo, B., et al. (2018). Insect-inspired neuromorphic computing. *Current Opinion in Insect Science*, 59–66.
De Andresbragado, L., & Sprecher, S. G. (2019). Mechanisms of vision in the fruit fly. *Current Opinion in Insect Science*, 25–32.
Eckert, H. (1980). Functional properties of the H1-neurone in the third optic ganglion of the blowfly, Phaenicia. *Journal of Comparative Physiology*, *135*(1), 29–39.
Eichner, H., Joesch, M., Schnell, B., et al. (2011). Internal structure of the fly elementary motion detector. *Neuron*, *70*(6), 1155–1164.
Fischbach, K. F., & Dittrich, A. P. M. (1989). The optic lobe of drosophila melanogaster. I. A Golgi analysis of wild-type structure. *Cell and Tissue Research*, *258*(3), 441–475.
Fox, J. L., & Frye, M. A. (2014). Figure-ground discrimination behavior in Drosophila. II. Visual influences on head movement behavior. *Journal of Fish Biology*, *217*(4), 570–579.
Fu, Q., Bellotto, N., Hu, C., et al. (2018). Performance of a visual fixation model in an autonomous micro robot inspired by drosophila physiology. In *2018 IEEE international conference on robotics and biomimetics (ROBIO)* (pp. 1802–1808). IEEE.
Fu, Q., Hu, C., Peng, J., et al. (2018). Shaping the collision selectivity in a looming sensitive neuron model with parallel ON and OFF pathways and spike frequency adaptation. *Neural Networks*, *106*, 127–143.
Fu, Q., Hu, C., Peng, J., et al. (2019). A robust collision perception visual neural network with specific selectivity to darker objects. *IEEE Transactions on Systems, Man, and Cybernetics*, 1–15.
Fu, Q. B., & Yue, S. G. (2015). Modelling LGMD2 visual neuron system. In *2015 IEEE 25th international workshop on machine learning for signal processing (MLSP)*. IEEE.
Hassenstein, B., & Reichardt, W. (1956). System theoretische analyse der zeit-, reihenfolgen-und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus. *Zeitschrift für Naturforschung B*, *11*(9–10), 513–524.
Hausen, K. (1976). Functional characterization and anatomical identification of motion sensitive neurons in the lobula plate of the blowfly Calliphora erythrocephala. *Zeitschrift für Naturforschung*, *31*(9–10), 629–634.
Hennig, P., & Egelhaaf, M. (2012). Neuronal encoding of object and distance information: a model simulation study on naturalistic optic flow processing. *Frontiers in Neural Circuits*, *6*, 14.
Hu, C., Arvin, F., Xiong, C., et al. (2017). Bio-inspired embedded vision system for autonomous micro-robots: the LGMD case. *IEEE Transactions on Cognitive and Developmental Systems*, *9*(3), 241–254.
Hu, C., Fu, Q. B., Yue, S. G., et al. (2018). Colias IV: The affordable micro robot platform with bio-inspired vision. In *19th annual conference towards autonomous robotic systems*. Cham: Springer.
Ito, K., Shinomiya, K., Ito, M., et al. (2014). A systematic nomenclature for the insect brain. *Neuron*, *81*(4), 755–765.
Joly, J. S., Recher, G., Brombin, A., et al. (2016). A conserved developmental mechanism builds complex visual systems in insects and vertebrates. *Current Biology*, *26*(20), R1001–R1009.
Keles, M. F., & Frye, M. A. (2017). Object-detecting neurons in drosophila. *Current Biology*, *27*(5), 680–687.
Kim, I. S., & Dickinson, M. H. (2017). Idiothetic path integration in the fruit fly Drosophila melanogaster. *Current Biology*, *27*(15), 2227–2238, e3.
Levine, D. S. (2018). *Introduction to neural and cognitive modeling* (third ed.). Routledge.
Liang, H., Morie, T., Suzuki, Y., et al. (2007). An FPGA-based collision warning system using hybrid approach. In *7th International conference on hybrid intelligent systems (HIS 2007)* (pp. 30–35). IEEE.
Lindemann, J. P., Kern, R., Van Hateren, J. H., et al. (2005). On the computations analyzing natural optic flow: quantitative model analysis of the blowfly motion vision pathway. *Journal of Neuroscience*, *25*(27), 6435–6448.
Longden, K. D., Wicklein, M., Hardcastle, B. J., et al. (2017). Spike burst coding of translatory optic flow and depth from motion in the fly visual system. *Current Biology*, *27*(21), 3225–3236, e3.
Ma, Z., & Krings, A. (2009). Insect sensory systems inspired computing and communications. *Ad Hoc Networks*, *7*(4), 742–755.
Maisak, M. S., Haag, J., Ammer, G., et al. (2013). A directional tuning map of Drosophila elementary motion detectors. *Nature*, *500*(7461), 212.
Mauss, A. S., & Borst, A. (2016). Electrophysiological recordings from lobula plate tangential cells in Drosophila. In *Drosophila* (pp. 321–332). New York, NY: Humana Press.
Mauss, A. S., Meier, M., Serbe, E., et al. (2014). Optogenetic and pharmacologic dissection of feedforward inhibition in drosophila motion vision. *The Journal of Neuroscience*, *34*(6), 2254–2263.
Missler, J. M., & Kamangar, F. A. (1995). A neural network for pursuit tracking inspired by the fly visual system. *Neural Networks*, *8*(3), 463–480.
Nasir, N., & Mat, S. (2019). An automated visual tracking measurement for quantifying wing and body motion of free-flying houseflies. *Measurement*, *143*, 267–275.
Oliva, D., & Tomsic, D. (2014). Computation of object approach by a system of visual motion-sensitive neurons in the crab neohelice. *Journal of Neurophysiology*, *112*(6), 1477–1490.
Omoto, J. J., Keles, M. F., Nguyen, B. M., et al. (2017). Visual input to the drosophila central complex by developmentally and functionally distinct neuronal populations. *Current Biology*, *27*(8), 1098–1110.
Pant, V., & Higgins, C. M. (2012). Tracking improves performance of biological collision avoidance models. *Biological Cybernetics*, *106*(4–5), 307–322.
Pfeiffer, K., & Homberg, U. (2014). Organization and functional roles of the central complex in the insect brain. *Annual Review of Entomology*, *59*, 165–184.
Pitonakova, L., & Bullock, S. (2020). The robustness-fidelity trade-off in grow when required neural networks performing continuous novelty detection. *Neural Networks*, 183–195.
Schneider, J., Murali, N., Taylor, G. W., et al. (2018). Can drosophila melanogaster tell who's who? *PLOS ONE*, *13*(10).
Sun, H., & Frost, B. J. (1998). Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons. *Nature Neuroscience*, *1*(4), 296.
Sun, H., Liu, L., & Guo, A. (1999). A neurocomputational model of figure-ground discrimination and target tracking. *IEEE Transactions on Neural Networks*, *10*(4), 860–884.
Taherkhani, A., Belatreche, A., Li, Y., et al. (2020). A review of learning in biologically plausible spiking neural networks. *Neural Networks*, 253–272.
Tammero, L. F., & Dickinson, M. H. (2002). Collision-avoidance and landing responses are mediated by separate pathways in the fruit fly, Drosophila melanogaster. *Journal of Fish Biology*, *205*(18), 2785–2798.
Temizer, I., Donovan, J. C., Baier, H., et al. (2015). A visual pathway for looming-evoked escape in larval zebrafish. *Current Biology*, *25*(14), 1823–1834.

Tuthill, J. C., Nern, A., Holtz, S. L., et al. (2013). Contributions of the 12 neuron classes in the fly lamina to motion vision. *Neuron*, *79*(1), 128–140.

Wang, H., Dewell, R. B., Zhu, Y., et al. (2018). Feedforward inhibition conveys time-varying stimulus information in a collision detection circuit. *Current Biology*, *28*(10), 1509–1521000.

Wang, H., Peng, J., Baxter, P., et al. (2018). A model for detection of angular velocity of image motion based on the temporal tuning of the drosophila. In *International conference on artificial neural networks* (pp. 37–46). Cham: Springer.

Wang, H. X., Peng, J. G., & Yue, S. G. (2017). An improved LPTC neural model for background motion direction estimation. In *2017 Joint IEEE international conference on development and learning and epigenetic robotics (ICDL-EpiRob)*. IEEE.

Weng, M., Huang, G., & Da, X. (2010). A new interframe difference algorithm for moving target detection. In *2010 3rd international congress on image and signal processing, Vol. 1* (pp. 285–289). IEEE.

Xiao, Q., Li, D. P., & Wang, S. R. (2006). Looming-sensitive responses and receptive field organization of telencephalic neurons in the pigeon. *Brain Research Bulletin*, *68*(5), 322–328.

Yue, S. G., & Rind, F. C. (2006a). Collision detection in complex dynamic scenes using an LGMD-based visual neural network with feature enhancement. *IEEE Transactions on Neural Networks*, *17*(3), 705–716.

Yue, S., & Rind, F. C. (2006b). Visual motion pattern extraction and fusion for collision detection in complex dynamic scenes. *Computer Vision and Image Understanding*, *104*(1), 48–60.

Yue, S., & Rind, F. C. (2013). Postsynaptic organisations of directional selective visual neural networks for collision detection. *Neurocomputing*, *103*, 50–62.

Zhang, Z., Yue, S. G., & Zhang, G. P. (2015). Fly visual system inspired artificial neural network for collision detection. *Neurocomputing*, *153*, 221–234.

Zufferey, J. C., & Floreano, D. (2006). Fly-inspired visual steering of an ultralight indoor aircraft. *IEEE Transactions on Robotics*, *22*(1), 137–146.