nature nanotechnology

Article

https://doi.org/10.1038/s41565-023-01379-2

Optoelectronic graded neurons for bioinspired in-sensor motion perception

In the format provided by the authors and unedited

This supplementary information includes:

Supplementary Note I to XI

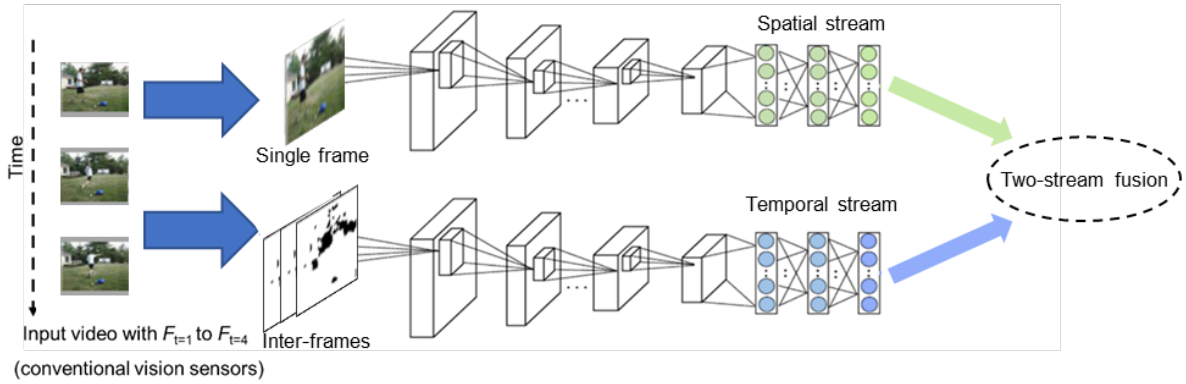
Supplementary Figure 1 to 20

Supplementary Table 1

Supplementary References

Supplementary Note I. Conventional vision chips with the complicated neural network for action recognition

Conventional CMOS image sensors collect static frames first. The spatial frames form a temporal video, further processed by the neural network for action recognition. Multilayer convolution neural network with spatial and temporal streams is a classical network for action recognition. The spatial stream performs action recognition from static frames, while the temporal stream is trained to recognize action from the motion information[1,2]. Each stream in this model is implemented with the multilayer convolution neural network. The scores of the two streams are combined via a late fusion to obtain the result for action recognition (**Supplementary Fig. 1**).



Supplementary Fig. 1| Action recognition based on the two-stream neural network. Processing of the dynamic visual signal in the two-stream computing model for action recognition. The spatial stream performs image recognition from static frames, while the temporal stream is trained to recognize action from the motion information. A late fusion combines the output neuron values of the two streams to realize action recognition.

Supplementary Note II. The temporal response of the spiking and graded neurons

Flying insects can acutely perceive human's motion with the graded neurons (**Supplementary Fig. 2**). Supplementary Table I compares the volatile characteristics of spiking and graded neurons. The volatile output potential in the spiking neurons can be described by Equation (1)[3]:

$$O(t) = \sum_f \eta(t - t^f) + \int_0^\infty \kappa(s) I_{sti}(t - s)\, ds + O_{rest} \qquad (1)$$

where $O(t)$ is the neuron output (membrane potential), O_{rest} is the resetting potential, $\eta(t - t^f)$ is refractory kernel, which describes the reset after the spike and the time course of the spike-afterpotential following a spike. t is the time, t^f stand for a spike at that time, $\kappa(s)$ is the leakage function of the spike input at time s, and I_{sti} is the stimulation. Based on Equation (1), **Supplementary Fig. 3b** shows the corresponding simulated results of the spiking neurons.

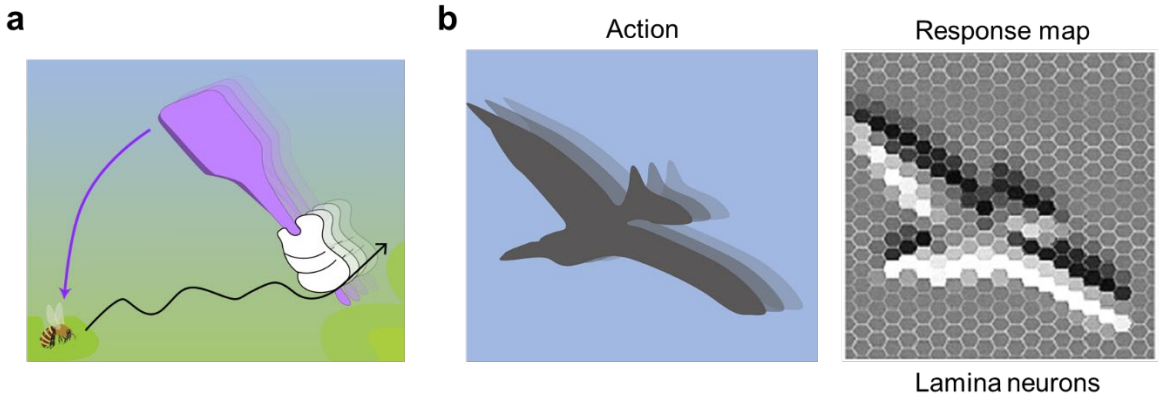
The volatile output potential in the graded neurons can be described by Equation (2) [3]:

$$\tau_m \frac{d}{dt} O(t) = E_L - O(t) + R I_{sti}(t) \qquad (2)$$

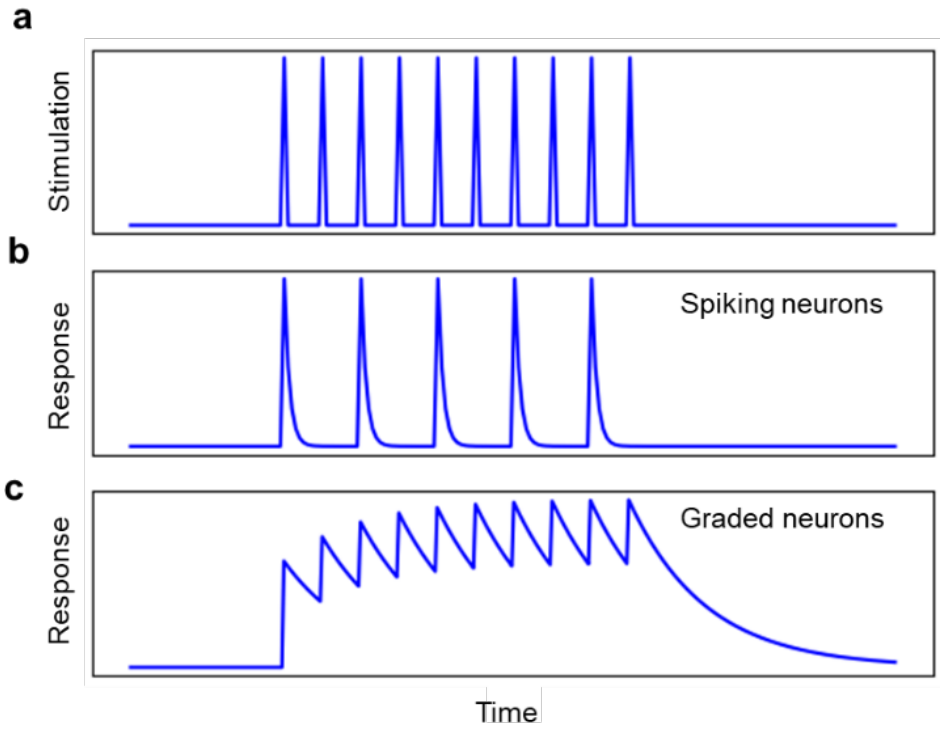
where τ_m is the membrane time constant, $O(t)$ is the neuron output (membrane potential), E_L is the leak potential, R is the membrane resistance, and $I_{sti}(t)$ is the stimulation. The simulated results (**Supplementary Fig. 3c)** based on Equation (2) agree well with the discrete and temporal summation responses to different sequential stimulation.

In the spiking neuron, a stimulation higher than the activation threshold can trigger depolarization and cause the action potential ("spike") to transmit information over long distances. A refractory period exists before the spiking neurons return to the resting potential, during which the neurons cannot encode the external stimulation. In the graded neurons, with the low-frequency input, the response increases and then decays quickly; with multiple inputs at a high frequency, the response is summed in the temporal domain and then decays to zero.

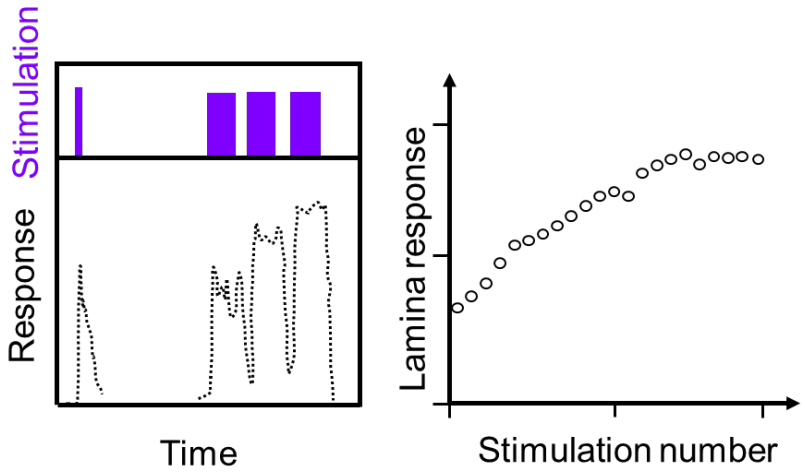
Supplementary Fig. 4 exhibits the nonlinear responses of the lamina neuron to repeated light stimulation (collected by the electrophysiological investigation[4,5]) due to the gradual increase in histamine-gated chloride conductance. In the visual system of insects, this nonlinear response characteristic can enhance temporal representation and increase the signal-to-noise ratio (*SNR*) of lamina neurons output, especially under dim light conditions or noisy environment[4].



Supplementary Fig. 2| Agile motion perception of insects. (a) Schematic illustration of a scene, in which a flying insect can agilely perceive human motion. (b) The mapping of lamina neuron arrays response to the motion. The light stimulation of present and past scene determines the output of lamina graded neurons. The response map is plotted from the calcium imaging microscopy study[6]. The lamina neuron arrays output the spatiotemporal information.



Supplementary Fig. 3| The responses of the spiking and graded neurons to the same input. (a) The stimulation as a function of time. (b) The response of spiking neurons. (c) The response of graded neurons



Supplementary Fig. 4| The responses of the lamina neuron to the optical stimulation. Figures are plotted based on the data in the electrophysiological studies[4,5].

Supplementary Table I| Comparison of spiking and graded neurons.

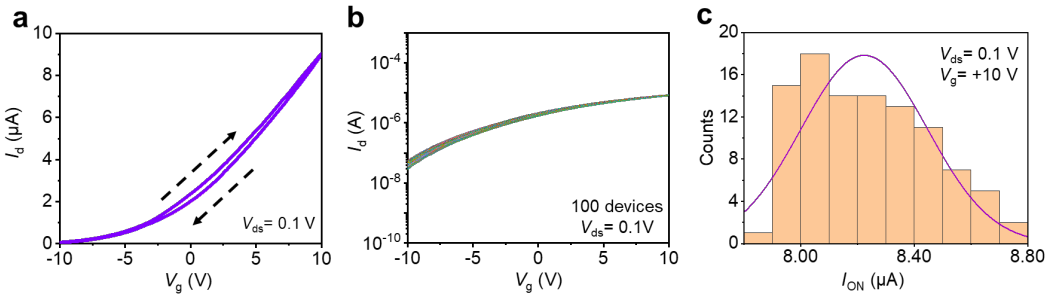
Spiking neuron	**Graded neuron**
Long axon	Short axon
Conducting action potentials	Conducting graded potentials
Pulsed transmitter release	Continuous transmitter release
Amplitude is all-or-none	Amplitude is proportional to the strength of the stimulus
Large amplitude	Amplitude is generally small
Absolute and relative refractory periods	No refractory period
Summation is not possible because of the all-or-none amplitude characteristic, and the presence of refractory periods	Graded potentials can be summed over time (temporal summation)

Supplementary Note III. The charge trapping and de-trapping processes in the MoS_2 transistor

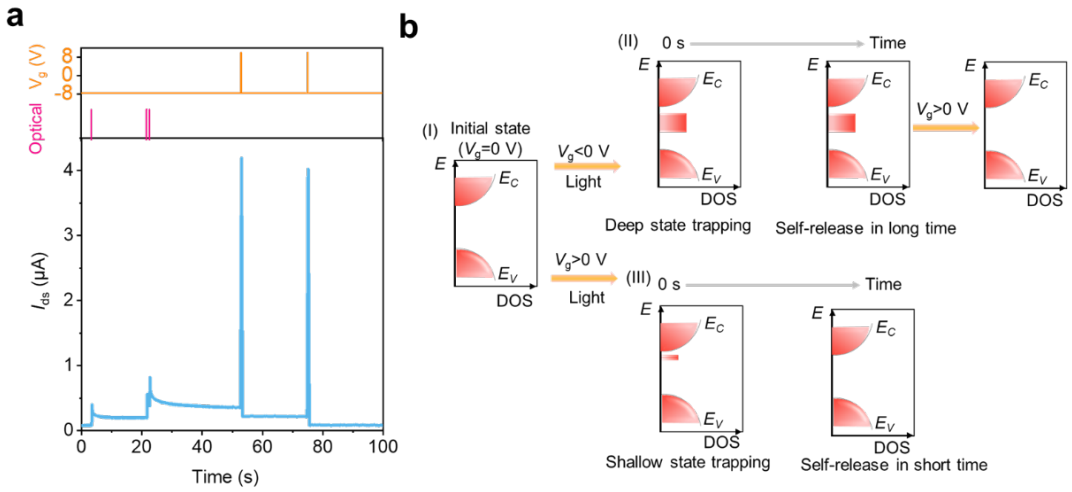
Supplementary Fig. 5a exhibits the typical transfer characteristic curves of the MoS_2 phototransistor under the dark condition. The observable clockwise hysteresis loop results support the charge trapping and de-trapping processes. The hysteresis voltage window (ΔV_{hys}) is defined as the difference in the gate voltage (V_g), which corresponds to $\Delta V_{hy} = \sim 1$V at the I_d of 2.2 μA. The trap charge density (N_t) is approximately 1×10^{12} cm^{-2} according to $N_t = (\Delta V_{hys} \times C_{ox})/q$, where C_{ox} is the oxide capacitance between the channel and local bottom gate, and q is the electron charge. The prepared devices show good uniformity (**Supplementary Fig. 5b** and **Supplementary Fig. 5c**).

The I_d is closely related to the optical stimulation and V_g (**Supplementary Fig. 6a**). Under V_g=-8 V and light pulses (10 mW/cm^2, 20 ms), the I_d increases fast and then decays slowly. After two sequential electrical pulses of V_g=+10 V (20 ms), the I_d resets gradually due to the sequential release of the trapped charges. The concentration of trapped charges affects the values of I_d. Larger trap concentration results in the higher I_d. **Supplementary Fig. 6b** illustrates the mechanism of gate-tunable charge trapping/de-trapping. Under negative voltages, deep trapping states exist in the MoS_2 under light illumination, which takes a long time for I_d to relax. Under positive voltages, the shallow trapping states enable fast relaxation. The controllable volatile characteristics provide the platform for processing temporal vision signals similar to the lamina neuron. To emulate the graded neurons, in our experiments, we mainly use the V_g of -3 to +3V and the response to the optical stimulation can volatilely drop to the baseline under these V_g.

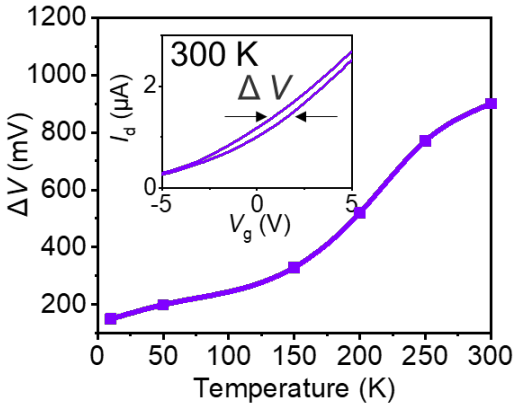
We conducted the tests of temperature-dependent transfer curves to analyze the concentration of trapping charges (**Supplementary Fig. 7**) from 10 K to 300 K. The hysteresis voltage window (ΔV_{hys}) is defined as the maximum difference in the transfer curve. As the temperature decreases, ΔV_{hys} becomes smaller, which is the typical experimental result of the charge trapping/de-trapping mechanism. Charge trapping/de-trapping is usually thermally activated, which is suppressed under the lower temperature. Our temperature-dependent characterization results can support the trapping/de-trapping mechanism of the devices.



Supplementary Fig. 5| Transfer curves of MoS_2 phototransistors with the Al_2O_3 encapsulation layer. (a) Typical drain current as a function of gating voltage of the MoS_2 phototransistor. (b) The transfer curves of 100 devices. V_{ds} is 0.1 V. With the Al_2O_3 encapsulation layer, the transfer curves show similar properties. (c) The statistical distribution of the on-state current (I_{on}) of 100 devices. V_g is +10 V.



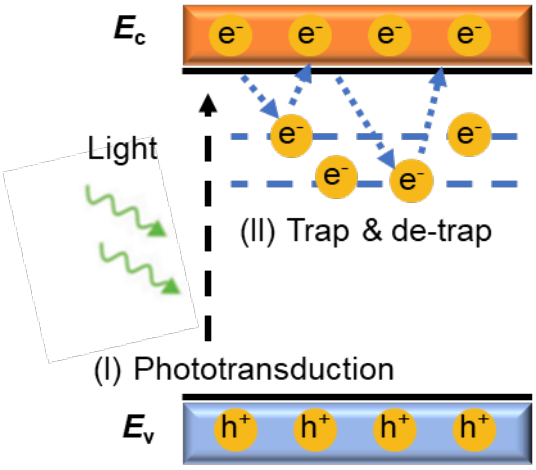
Supplementary Fig. 6| Gate-tunable charge trapping/de-trapping characteristics of MoS_2 phototransistors. (a) Positive and negative V_g on the I_d of MoS_2 phototransistors. (b) Mechanism of gate-tunable charge trapping/de-trapping.



Supplementary Fig. 7| Temperature-dependent transfer curves of the MoS_2 phototransistor.

Supplementary Note IV. MoS_2 phototransistor for emulating the elementary functions of the insect vision system

The first two layers of the insect vision system are the retina and lamina. The main role of the retina is the photoreceptor, acting as phototransduction. Light stimulation can change the graded membrane potential of the retina photoreceptor and release histamine, which transfers to the lamina neuron (a typical graded neuron) for temporal processing (**Fig. 1b**). Our optoelectronic devices emulate the phototransduction function of the retina by the interaction between light and MoS_2 channel (stage I in **Fig. Supplementary Fig. 8**). Then, the release of histamine to the lamina neuron is emulated through the release of photogenerated carriers to the trapping states of MoS_2 films (stage II in **Fig. Supplementary Fig. 8**).



Supplementary Fig. 8| The phototransduction and charge trapping processes in the MoS2 phototransistor.

Supplementary Note V. The time constant of the decaying I_d curve of phototransistors.

After the removal of light illumination, I_d gradually decreases because of the release of charges in the trap states, which can be described by Equation (3):

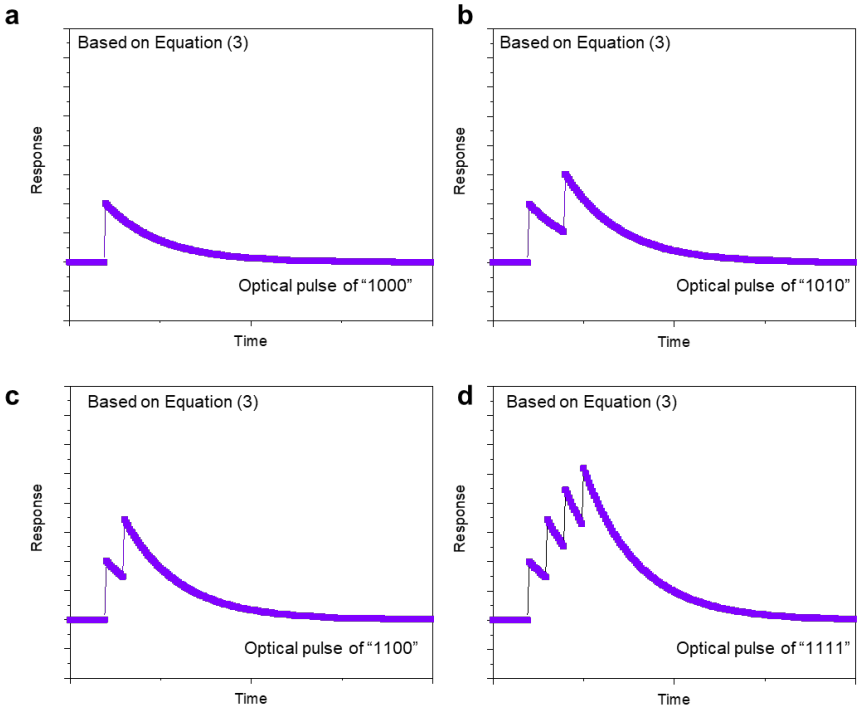
$$I(t) = I_{dark} + \left(I_{peak} - I_{dark}\right) \exp\left(-\frac{t}{t_0}\right) \quad (3)$$

where I_{dark} is the current under dark condition, I_{peak} is the maximum current upon light stimulation, t is the time, and t_0 is the time constant. Based on Equation (2), the extracted t_0 is ~100 ms for the I_d curve after the first light stimulation (**Fig. 2b**).

By combing Equation (2), we can modify the Equation (1) as the following Equation (4):

$$I(t, t_p) = I_{dark} + \int_0^t e^{-(t-t_p)/t_0} \times (I_{peak} - I_{dark}) \times \delta(t_p) dt_p \quad (4)$$

where I_{dark} is the dark current, t is the time, t_0 is the time constant, and t_p is the time step. **Supplementary Fig. 9** shows the typical simulated results based on Equation (4).



Supplementary Fig. 9| Simulated responses to typical stimulations based on Equation (4). (a) Response to the stimulation of "1000". (b) Response to the stimulation of "1010". (c) Response to the stimulation of "1100". (d) Response to the stimulation of "1111". "1" indicates applying the optical pulse while "0" indicates the dark condition.

Supplementary Note VI. Calculation details of the signal-to-noise ratio and flicker fusion frequency

Signal-to-noise ratio (SNR) has been widely used in science and engineering to compare the level of a desired signal to background noise. *SNR*(*f*) is the ratio between the signal power spectra ($|S(f)|^2$) and noise power spectra ($|N(f)|^2$) under the specific frequency by Equation (5)[5,7]:

$$SNR(f) = \frac{|S(f)|^2}{|N(f)|^2} \quad (5)$$

where $|S(f)|^2$ is extracted from the Fourier transform on the average photoresponse curve, and $|N(f)|^2$ is extracted from the Fourier transform on the corresponding noise traces (the differences between individual response and signal). Similar to the biological work[7], the tested data are windowed with a Blackman-Harris 4-term window before calculating $|S(f)|^2$ and $|N(f)|^2$. In the photoresponse curve, we adopt the identical light pulses to generate the output photocurrents. We describe the detailed calculation process in Equation (6) - (10). The average photoresponse can be calculated by Equation (6)

$$S(t) = \frac{S_1(t) + S_2(t) + \cdots + S_5(t)}{5} \tag{6}$$

where $S_1(t)$, $S_2(t)$, ..., $S_5(t)$ refer to the photocurrent response in the time domain with the 5 identical light stimuli. The noise trace for each photoresponse can be calculated by Equation (7):

$$N_n(t) = S_n(t) - S(t) \tag{7}$$

where $N_n(t)$ represents the noise trace under different stimulation cycles. The average photoresponse signal and noise traces are calculated with Fourier transformation with Equation (8) and Equation(9):

$$S(f) = \int_{-\infty}^{\infty} S(t) e^{-2\pi i t f} dt \tag{8}$$

$$N(f) = \int_{-\infty}^{\infty} N(t) e^{-2\pi i t f} dt \tag{9}$$

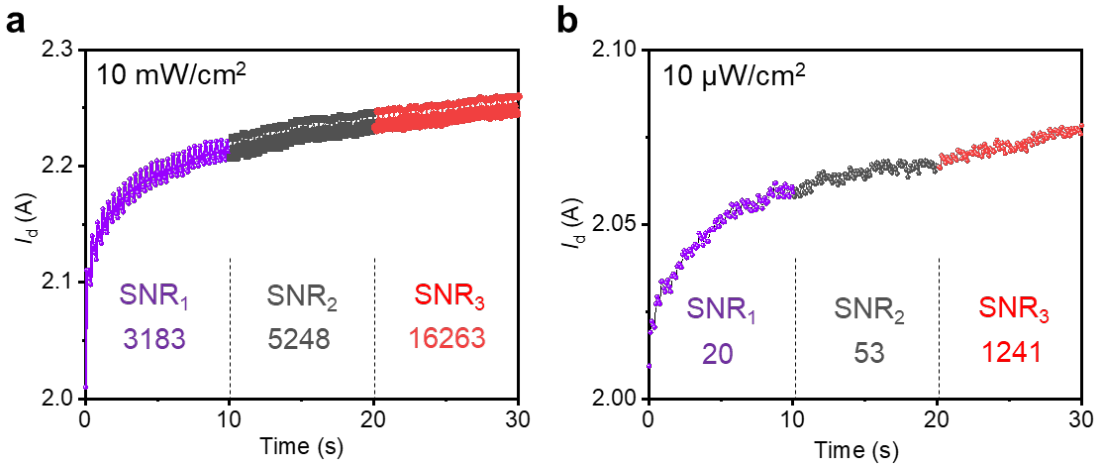
where f represent the frequency of the signal and noise.

As the signal-to-noise ratio (*SNR*) of 1 corresponds to the indistinguishable response between the light stimulation and the dark conditions, we use *SNR* =1 to extract the flicker fusion frequency (*FFF*).

Supplementary Note VII. Nonlinear response increases the *SNR*

The *SNR* has been widely used as the statistical analysis for comparing the levels of desired signal and background noise. We divided the nonlinear photoresponse curves under different

frequencies into three segments (**Supplementary Fig. 10**) and calculated the *SNR* of each segment (marked as SNR_1, SNR_2 and SNR_3). Generally, the *SNR* is stronger under the higher light intensity due to the more stable photoresposne. Under the light intensity of 10 mW/cm^2, we can acquire SNR_1=3183, SNR_2=5248 and SNR_3=16263; under the light intensity of 10 μW/cm^2, SNR_1=20, SNR_2=53 and SNR_3=1241. For the dim light (10 μW/cm^2), the SNR_3 is ~60 times higher than the SNR_1. These results show SNR_3> SNR_2> SNR_1 for both light intensities. Higher light intensity will incur higher photoresponse signal than the low intensity, while the noise level is similar. Thus, the *SNR* is larger for the higher light intensity.



Supplementary Fig. 10 | Light-intensity-dependent response to the light stimulation and *SNR* of three segments (marked as SNR_1, SNR_2 and SNR_3). (a) I_d and SNR under 10 mW/cm^2 with the duty ratio of 25%. (b) I_d and SNR under 10 μW/cm^2 with the duty ratio of 25%. Similar to the biological investigation of lamina neurons[4], we divide the curves into three parts and calculated the SNR of each segment.

Supplementary Note VIII. Calculation of information transmission rate of the phototransistors

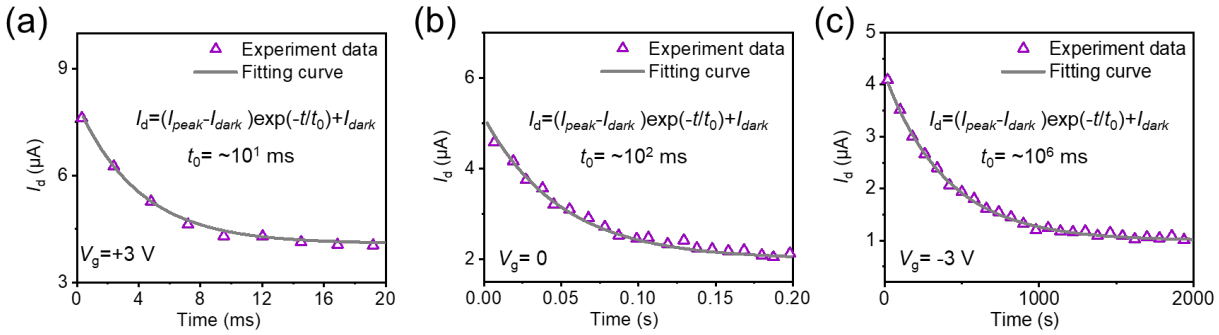
To estimate the information capacities of neurons, it is necessary to measure the reliability and dynamics of the neural coding process[8]. Information transmission rate (R) is a parameter for characterizing the amount of information transmission per unit time (bits per second)[9]. We can calculate the R based on $SNR(f)$ by Equation (11)[9]:

$$R = \int_1^{100} df\ log_2[1 + SNR(f)] \quad (11)$$

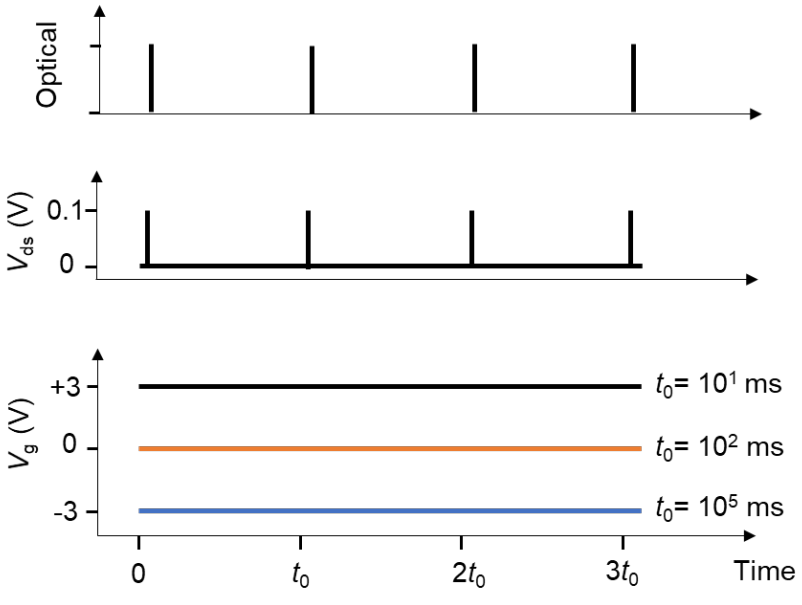
The calculated R is about 1200 bit/s when the frequency varies from 1 to 100 Hz under the light intensity of 10 mW/cm^2. For comparison, the graded neurons in the biological work[9] exhibit 1650 bit/s when the light stimulation frequency reaches about 500 Hz, and ~1000 bit/s when the light stimulation frequency varies from 1 to 100 Hz. Notably, the spiking neurons show the R of only ~300 bit/s[8].

Supplementary Note IX. Gate-tunable t_0 and corresponding sampling frequency

The t_0 of I_d varies from 10^1 to 10^6 ms when the V_g decreases from +3 V to -3 V (Supplementary Fig. 11). As the sampling frequency is set as $1/t_0$, the t_0 of 10 ms, 10^2 ms, and 10^6 ms correspond to the sampling frequency of 10^2 Hz, 10^1 Hz, and 10^{-3} Hz (Supplementary Fig. 12), respectively.



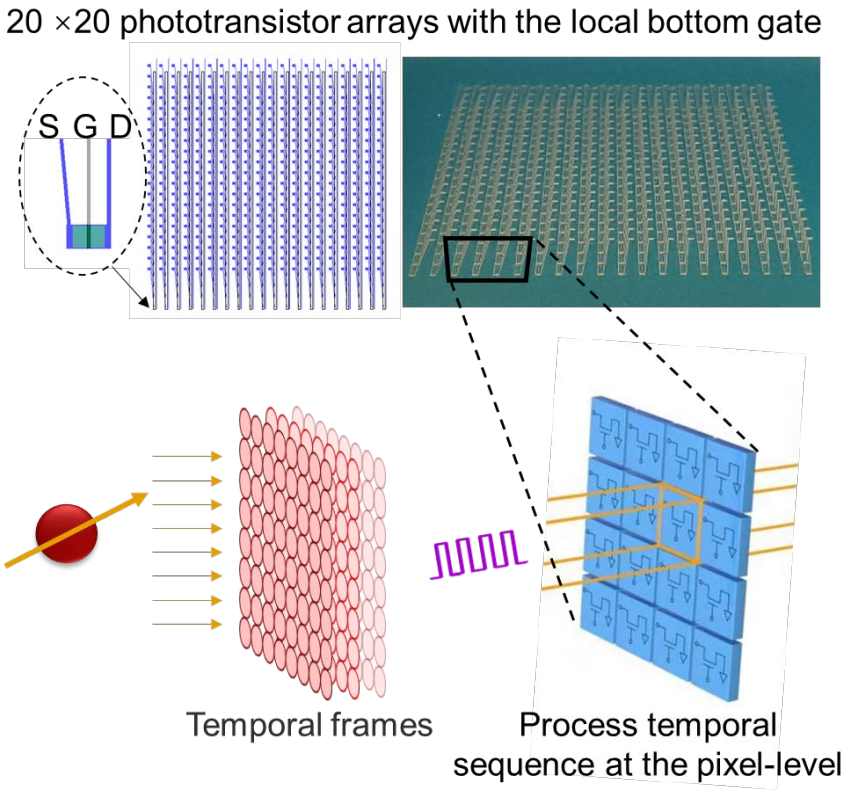
Supplementary Fig. 11| Gate-tunable decaying curve and time constant of I_d after the light stimulation. (a) I_d after the light stimulation under V_g= +3 V. The fitted time constant is ~10^1 ms. (b) I_d after the light stimulation under V_g= 0. The fitted time constant is ~10^2 ms. (c) I_d after the light stimulation under V_g= -3 V. The fitted time constant is ~10^6 ms.



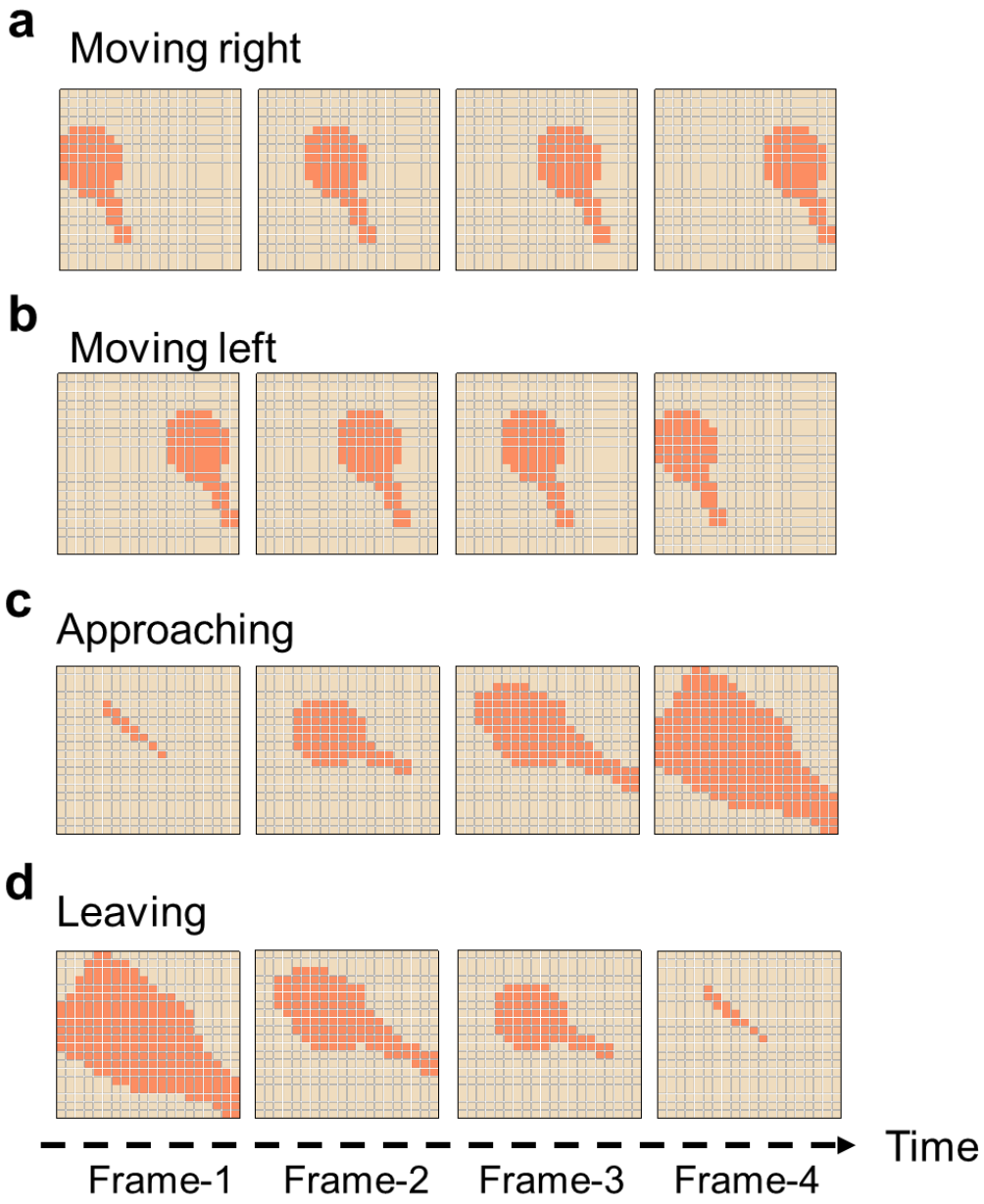
Supplementary Fig. 12| Operation of MoS_2 phototransistors for motion perception. The framerate (the frequency of the optical stimulation and readout V_{ds}) is set as $1/t_0$, which depends on the V_g.

Supplementary Note X. In-sensor motion perception with the photosensor array

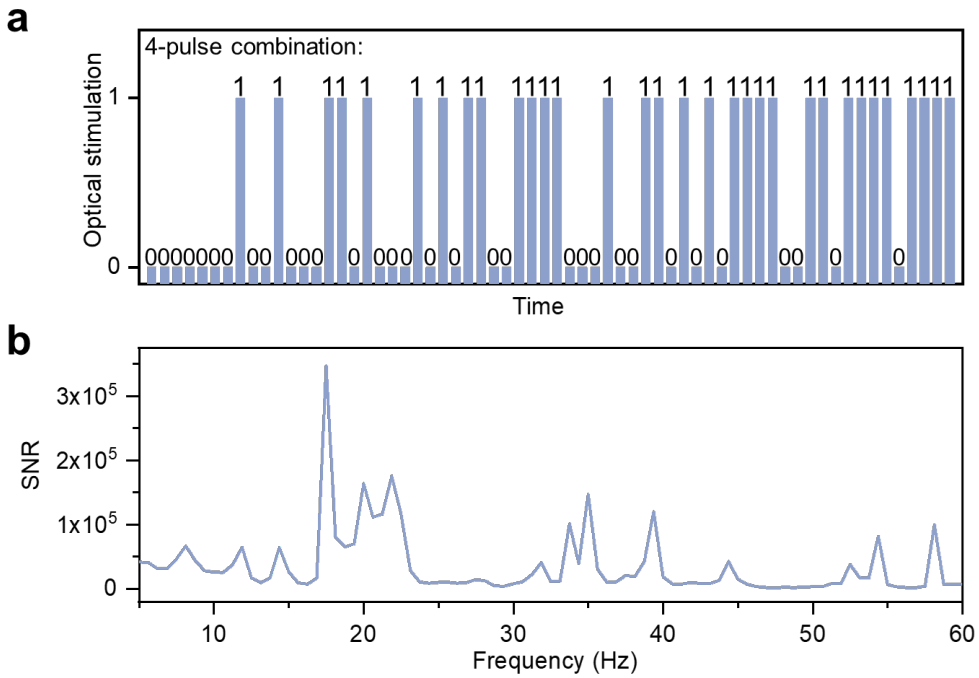
We adopt 20×20 photosensor array for encoding the spatiotemporal vision information (**Supplementary Fig. 13**). **Supplementary Fig. 14** exhibits the temporal evolution of visual stimulation on 20×20 bioinspired phototransistor array. The *SNR* of the responding curves to 4 optical pulses is high during the tested frequency **(Supplementary Fig. 15)**. The output I_d map **(Supplementary Fig. 16)** based on 20×20 bioinspired phototransistor array shows the contour of the right-moving, left-moving, the approaching and the leaving motion. **Supplementary Fig. 17** shows the target salience based on the bioinspired vision sensors.



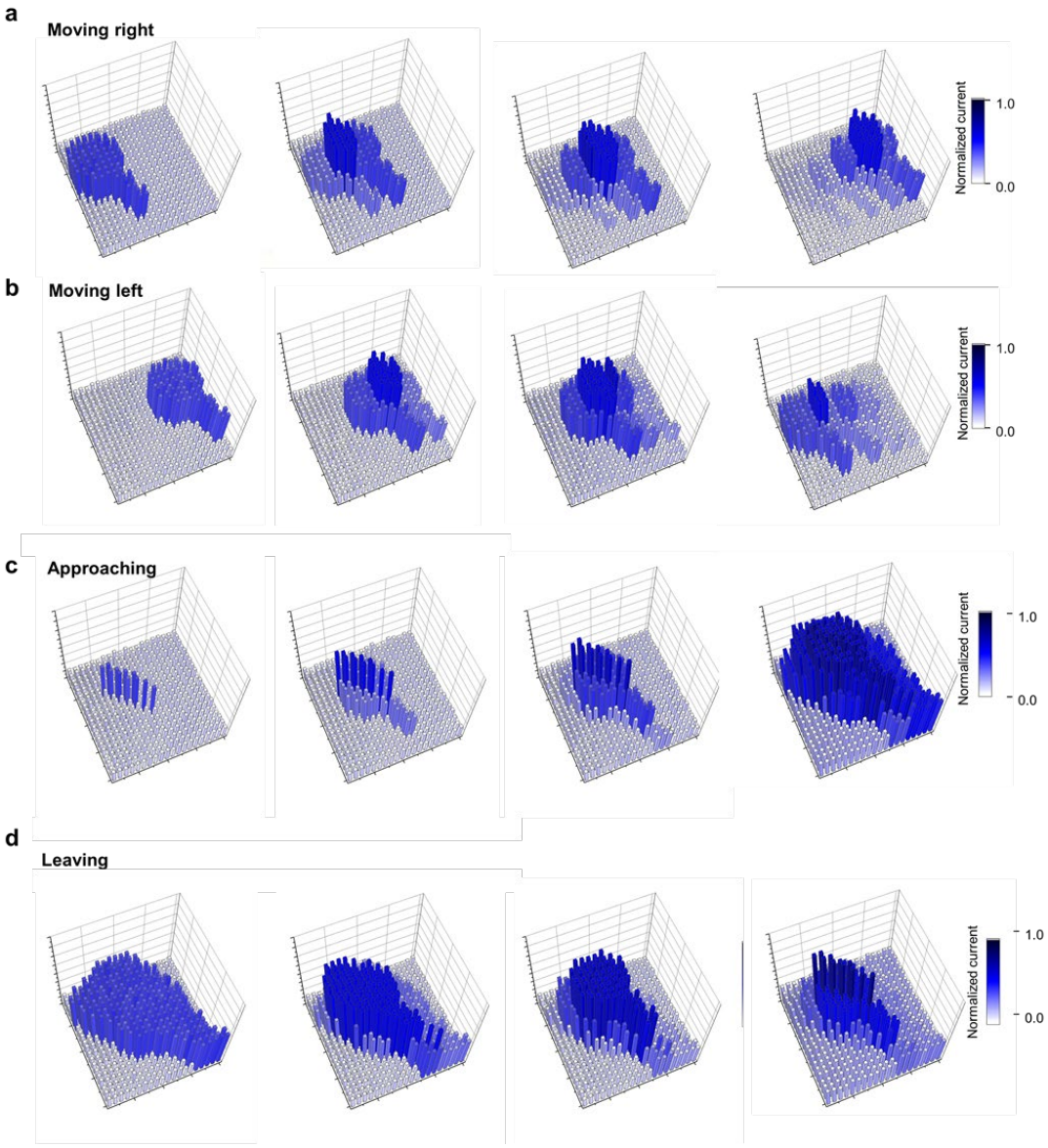
Supplementary Fig. 13| Perception of the temporal action with 20 ×20 phototransistor array. Each device with the local bottom gate structure can process the pixel sequence, formed by the temporal frames at the pixel level.



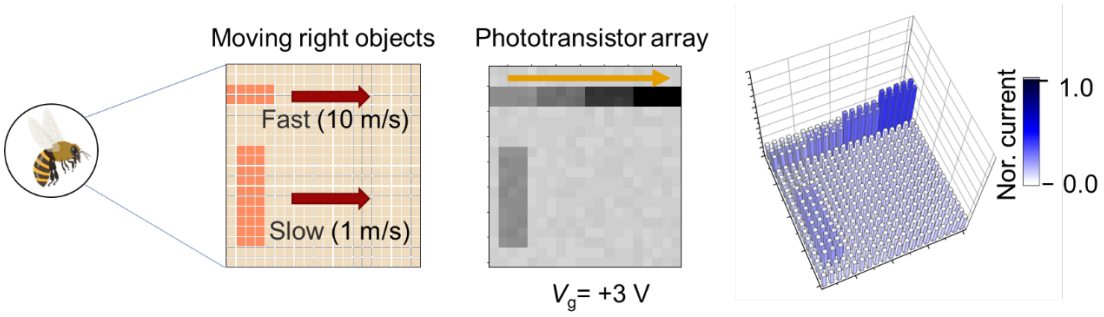
Supplementary Fig. 14| The temporal evolution of visual stimulation is projected on 20×20 bioinspired phototransistor array. (a) Temporal frames of a right-moving object. (b) Temporal frames of a left-moving object. (c) Temporal frames of the approaching motion. (d) Temporal frames of the leaving motion.



Supplementary Fig. 15| *SNR* of the responding curves to 4 light pulses. (a) 4-pulse stimulations as a function of time. (b) The calculated *SNR* as a function of the frequency.



Supplementary Fig. 16| The output I_d map based on 20×20 bioinspired phototransistor array. (a) I_d map for the temporal frames of a right-moving object. (b) I_d map for the temporal frames of a left-moving object. (c) I_d map for the temporal frames of the approaching motion. (d) I_d map for the temporal frames of the leaving motion.



Supplementary Fig. 17| Target salience based on the bioinspired vision sensors. There are slow-moving (0.5 m/s) with a duration time of 0.2 s and fast-moving objects (5 m/s) with a duration time of 0.02 s in the scene. There is only the trajectory contour of a fast-moving object.

Supplementary Note XI. Neuromorphic computing network for action recognition

The training and inference of vision systems are based on bioinspired vision sensor and conventional image sensors. The temporal frames of specific action are taken as discrete events. The value of each pixel in the temporal frame is converted into the light intensity. For the bioinspired vision sensors, the responses to the light intensity are based on the output in **Fig. 2e**. The response of conventional image sensors is only determined by the present stimulation.. Both systems are trained round by round. The accuracy is the testing result per round. The neuromorphic recognition result (**Fig. 4c**) shows that the system implemented with bioinspired vision sensors can recognize the actions with nearly 100% accuracy.

We investigated the effect of device-to-device variation on the recognition accuracy of the motion perception. The neuromorphic computing with the artificial neural network allows to tolerate a certain level of device variations because of its non-precise-calculation characteristics. During the stimulation, the device-to-device variation exist in both the training and testing stages. In the task of recognizing moving objects. **Supplementary Fig. 18** shows the recognition accuracy with different device variations. Even for the 25% device-to-device variation, the recognition accuracy after the training is still higher than 90%.

The trained vision systems can continuously monitor a typical sample of the complete left-moving process in the real world. In the last layer of the neural network, there are five output neurons ("up", "down", "right", "left", and "blank") with different values to specific action

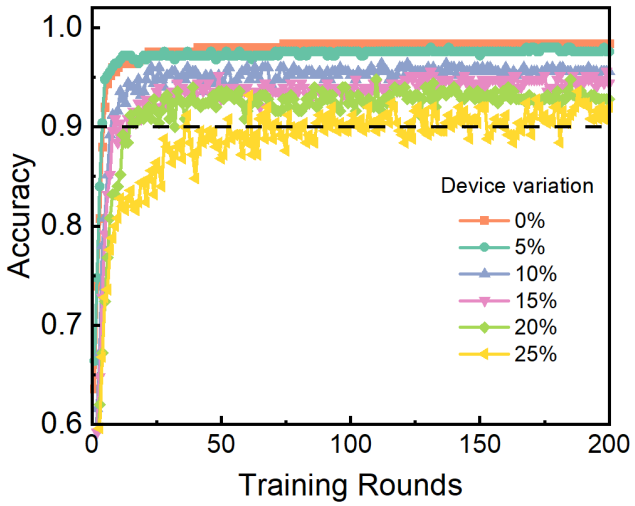
frames. The recognition of "moving left" is exported when the output neuron value of "left" meets the following two conditions: (1) the highest among five output neurons; (2) higher than the threshold value of 0.3. The recognition results (**Supplementary Fig. 19** and **Supplementary video 2**) reveal that the system based on conventional image sensors cannot recognize the left-moving ball. However, the system with our bioinspired vision sensors can efficiently recognize real-world action.

The bioinspired vision sensors can recognize moving objects with different speeds by adjusting the time constant of MoS_2 phototransistors. As shown in **Fig. 2f**, the MoS_2 phototransistors with different time constants can process the 4-pulse stimulation by adjusting the gate voltage. The time interval of inter-frames perceived by the MoS_2 phototransistor is set as the time constant values.

A well-trained vision system in fixed time constant is transferred to monitor the motion of solid balls with different speeds. **Supplementary Fig. 20** shows that the system monitors the same motion process with different time constants. The vision system with our bioinspired phototransistors can adjust its action recognition sensitivity for the objects with various speeds, which satisfies most action recognition scenes in the real world. While the vision sensors based on the threshold characteristics can increase the image contrast for high-accuracy static image recognition[10], the optoelectronic devices with the graded response characteristics show the capability of processing dynamic motion to realize action recognition.

The dataset of the moving ball in this study for action recognition tasks is available at https://github.com/ZhengZhouPKU/Moving_Ball_Dataset.

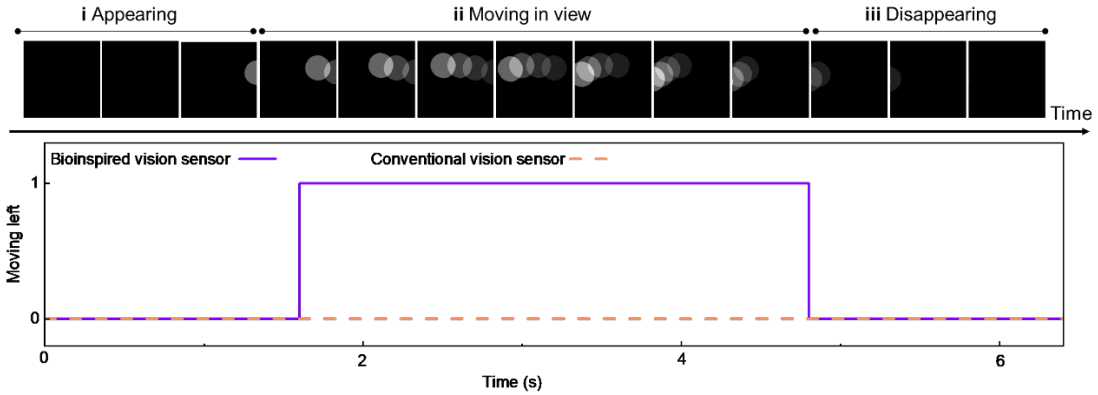
The samples of the real-world wild animal's actions with various timescale are available at https://github.com/ZhengZhouPKU/Wild_Animal_Motion_Samples.



298

299 **Supplementary Fig. 18| The effects of device-to-device variation on the recognition**
300 **accuracy.** The device variation varies from 0% to 25%.

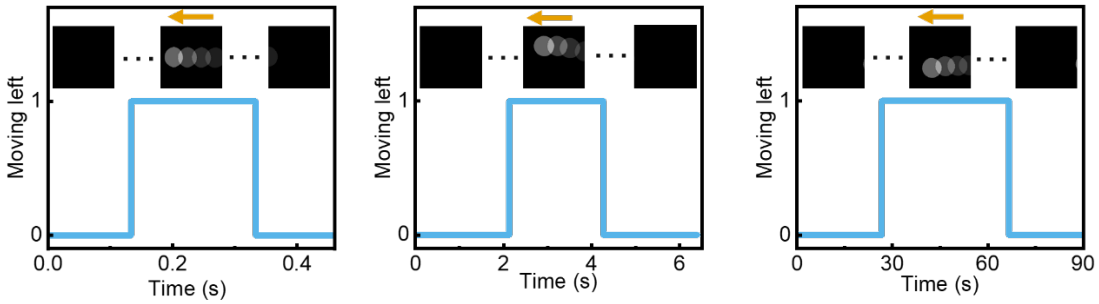
301



302

303 **Supplementary Fig. 19| Monitor a moving-left ball with the neural network based on**
304 **bioinspired and conventional vision sensors, respectively.** Top panel: temporal frames of
305 specific left-moving action, including the appearing, moving in view and disappearing
306 processes. Bottom panel: action recognition as time increases. Based on bioinspired vision
307 sensors, the neural network can accurately recognize the left-moving action when the ball
308 moves in view.

309



Supplementary Fig. 20| Action recognition of a left-moving ball with different duration time. The inset shows the temporal frames of a left-moving ball. Based on the bioinspired vision sensors, the neural network can accurately output the recognition result of left motion (represented by "1").

Supplementary References

1 Yao, G., Lei, T. & Zhong, J. A review of convolutional-neural-network-based action recognition. *Pattern Recognition Letters* **118**, 14-22 (2019).

2 Ye, H. *et al.* in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval.* 435-442.

3 Gerstner, W., Kistler, W. M., Naud, R. & Paninski, L. *Neuronal dynamics: From single neurons to networks and models of cognition*. (Cambridge University Press, 2014).

4 Zheng, L. *et al.* Network adaptation improves temporal representation of naturalistic stimuli in Drosophila eye: I dynamics. *PLoS One* **4**, e4307 (2009).

5 Nikolaev, A. *et al.* Network adaptation improves temporal representation of naturalistic stimuli in Drosophila eye: II mechanisms. *PloS one* **4**, e4306 (2009).

6 Schuetzenberger, A. & Borst, A. Seeing Natural Images through the Eye of a Fly with Remote Focusing Two-Photon Microscopy. *Iscience* **23**, 101170 (2020).

7 Juusola, M. & Hardie, R. C. Light adaptation in Drosophila photoreceptors: I. Response dynamics and signaling efficiency at 25 C. *The Journal of general physiology* **117**, 3-25 (2001).

8 Laughlin, S. B., de Ruyter van Steveninck, R. R. & Anderson, J. C. The metabolic cost of neural information. *Nature Neuroscience* **1**, 36-41 (1998).

9 De Ruyter van Steveninck, R. & Laughlin, S. The rate of information transfer at graded-potential synapses. *Nature* **379**, 642-645 (1996).

10 Seung, H. *et al.* Integration of synaptic phototransistors and quantum dot light-emitting diodes for visualization and recognition of UV patterns. *Science Advances* **8**, eabq3101 (2022).