















Ultrafast visual perception beyond human capabilities enabled by motion analysis using synaptic transistors

Received: 16 April 2025

Accepted: 12 January 2026

Published online: 10 February 2026

 Check for updates

Shengbo Wang ^{1,2,14,15}, Jingwen Zhao ^{3,4,15}, Tongming Pu ^{2,15}, Liangbing Zhao ^{5,15}, Xiaoyu Guo ⁶ ✉, Yue Cheng ^{3,4}, Cong Li², Weihao Ma², Chenyu Tang ⁷, Zhenyu Xu^{2,8}, Ningli Wang ⁹, Luigi G. Occhipinti ⁷, Arokia Nathan ^{10,11}, Ravinder Dahiya ¹², Huaqiang Wu ¹³, Li Tao ^{3,4} ✉ & Shuo Gao ^{1,2} ✉

Optical flow, inspired by biological visual systems, calculates spatial motion vectors aiming to enable robotics to excel in dynamic environments. However, current algorithms, despite human-competitive task performance on benchmark datasets, suffer from significant time delays, limiting practical deployment. Here, we introduce a neuromorphic temporal-attention hardware that emulates the interaction between the retina and the lateral geniculate nucleus (LGN) to extract temporal motion cues directly in hardware. Using a two-dimensional synaptic transistor array, the system encodes brightness changes and accumulates them in analog, non-volatile states, generating compact regions of interest (ROIs). These ROIs then act as inputs to conventional downstream optical flow and vision algorithms, enabling ultrafast motion analysis. At the hardware level, the synaptic transistor offers high-frequency response (~100 μ s), non-volatility (>10,000 s), and endurance (>8,000 cycles). Compared to state-of-the-art algorithms, our approach demonstrates a 400% speedup, surpassing human-level performance while maintaining or improving accuracy through temporal priors.

Optical flow, originally introduced by James J. Gibson in the 1950s and inspired by biological visual perception, estimates motion vectors within a visual scene^{1–3}. Over the decades of development, current state-of-the-art algorithms such as RAFT and GMFlow have

demonstrated impressive performance on benchmark datasets. By leveraging two-dimensional gradient representations of pixel movement, optical flow offers clear, intermediate motion representations compared to end-to-end approaches, enabling it to excel in dynamic

¹Hangzhou International Innovation Institute, Beihang University, Hangzhou, China. ²School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China. ³Center for Quantum Physics, Key Laboratory of Advanced Optoelectronic Quantum Architecture and Measurement (MOE), School of Physics, Beijing Institute of Technology, Beijing, China. ⁴State Key Laboratory of Chips and Systems for Advanced Light Field Display, Center for Interdisciplinary Science of Optical Quantum and NEMS Integration, School of Physics, Beijing Institute of Technology, Beijing, China. ⁵Computer Science, Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. ⁶Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, China. ⁷Department of Engineering, University of Cambridge, Cambridge, UK. ⁸Department of Precision Instrument, Tsinghua University, Beijing, China. ⁹Beijing Tongren Hospital, Capital Medical University, Beijing, China. ¹⁰Darwin College, University of Cambridge, Cambridge, UK. ¹¹School of Information Science and Engineering, Shandong University, Qingdao, China. ¹²Bendable Electronics and Sustainable Technologies (BEST) Group, Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. ¹³School of Integrated Circuits, Tsinghua University, Beijing, China. ¹⁴Present address: Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China. ¹⁵These authors contributed equally: Shengbo Wang, Jingwen Zhao, Tongming Pu, Liangbing Zhao.

✉ e-mail: xiaoyuguo@cityu.edu.hk; litaob@bit.edu.cn; shuo_gao@buaa.edu.cn

scene analysis at performance levels comparable to human capabilities^{4–9}. However, translating these achievements to real-world applications is still challenging due to the high computational overhead required to process visual inputs in real time. For instance, while Tesla’s Autopilot employs occupancy networks to achieve latency as low as ~ 10 ms^{10,11}, performing optical flow analysis and object segmentation on a 1920×1080 resolution image can require over 0.6 s on an Nvidia V100 GPU, 4X humans (Supplementary Notes 1, 2). Such delays are unacceptable for time-sensitive applications like autonomous driving, where a one-second delay at highway speeds can reduce the safety margin by up to 27 m, significantly increasing safety risks. Thus, due to this obvious time delay, the field deployment of optical flow seems to be unrealistic^{12–15}.

Optical flow is initially aimed to mimic the processing pipeline of biological visual systems, but cannot faithfully replicate their high processing efficiency in practice, which is because biological vision excels at processing large volumes of visual information efficiently through dynamically focusing on regions where motion occurs^{16,17}. Specifically, biological vision, including the retina and lateral geniculate nucleus (LGN), emphasizes temporal change and dynamically routes information toward locations where motion occurs, acting as a temporal-attention function that detects moving regions before high-level complex processing^{18,19}. This suggests a design principle for artificial vision: a fast, low-overhead function that dynamically detects motion regions, thereby accelerating downstream motion analysis by focusing computation where changes occur. Replicating this dynamic processing in artificial systems is challenging due to the intrinsic inflexibility of conventional CMOS-based technologies, which cannot readily adjust their processing functions in response to varying stimuli^{20,21}. Fortunately, neuromorphic devices such as synaptic transistors and memristors offer synapse-like characteristics that can emulate the processing functions of biological synapses. Their intrinsic adaptability enables on-device adaptation and local feature extraction functions, which in turn lowers both energy consumption and processing latency. This approach has already been applied across multiple sensory modalities. In a typical implementation, a front-end sensor first transduces external stimuli into electrical signals (voltage or current), which are then transmitted to synaptic devices for processing. For tactile perception, the sensor is typically a piezoresistive

or piezoelectric film that converts mechanical force into an electrical response. Based on this architecture, artificial tactile receptors are widely studied—including rapidly adapting mechanoreceptors and nociceptors—using neuromorphic hardware^{22–25}. For example, Wang et al. proposed a memristor-based perceptual signal processing method, which can emulate multiple essential tactile receptors and sensory neurons with a single memristor²². For auditory processing, neuromorphic device arrays can capture and process auditory temporal patterns directly in hardware, enabling low-latency, low-power front ends for sound recognition and localization^{26–28}. For visual processing, the high spatial redundancy of video frames aligns perfectly with the local, in-memory processing capability of synaptic devices^{22,29–31}. For example, Baek et al. recently reported a neuromorphic neuron with dendritic morphology based on silicon-nanowire transistors that can perform nonlinear integration of excitatory/inhibitory synaptic inputs while considering spatial distribution dependency. Its built-in direction selectivity enables on-device motion detection, and the power consumption of processing event-based pulses from the 1000 pixels is about 0.1–2.0 mW³². More broadly, the on-device adaptation capability of synaptic devices makes human-like ultrafast visual processing possible^{22,33–35}.

Here, inspired by the biological LGN system, we propose a neuromorphic hardware that directly detects motion changes using two-dimensional neuromorphic synaptic transistors. The array encodes brightness derivatives and accumulates these temporal changes in analog, non-volatile states, as shown in Fig. 1. The resulting array state generates temporal motion cues, i.e., compact maps of regions of interest (ROIs) that guide and accelerate downstream optical flow calculation. In our demonstration, conventional optical flow calculation methods (e.g., Farneback, GMFlow, and RAFT) are applied only inside these ROIs, not the entire image, ultimately accelerating various tasks such as motion prediction, segmentation, and tracking. In summary, the neuromorphic hardware rapidly extracts motion change information, which is then passed to conventional processors for downstream optical flow computation in our demonstration (Supplementary Movie 1). At the device level, leveraging the superior properties of two-dimensional materials, such as atomic thickness and enhanced electrostatic control, the developed floating gate synaptic transistors demonstrate high-frequency response (~ 100 μ s), robust

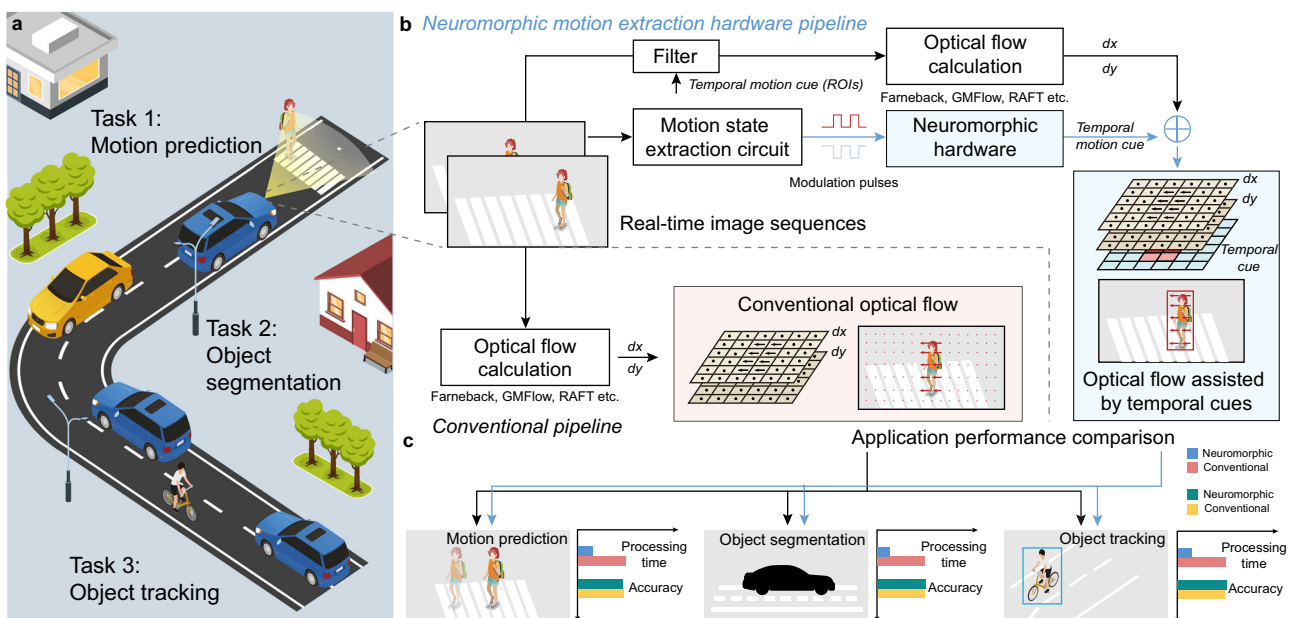


Fig. 1 | Neuromorphic motion extraction hardware and its application. **a** Example application scenarios. **b** Comparison between the proposed neuromorphic motion extraction hardware pipeline and the conventional optical flow pipeline. **c** Performance comparison of processing time and accuracy between the two approaches.

retention ($>10^4$ s), and exceptional endurance (>8000 cycles). In our experiments, we deployed our pipelines across various application scenarios—including vehicle operation, UAVs, robotic arms, and sports activities—to perform tasks like motion prediction, object segmentation, and object tracking. The results demonstrate that our method significantly accelerates processing times, achieving an average of $\sim 400\%$ speedup and surpassing human-level speeds (~ 150 ms) in most cases. Notably, by incorporating spatial-temporal consistency of motion information, our spatiotemporal approach maintains or improves accuracy, such as a 213.5% performance increase in the vehicle scenario. These advancements empower robots with ultrafast and accurate perceptual capabilities, enabling them to handle complex and dynamic tasks more efficiently than ever before.

Results

Floating gate synaptic transistor

In our neuromorphic motion extraction hardware, synapse arrays serve to embed temporal information from external visual scenes. To achieve precise encoding and long-term retention of this information, the synapse array must exhibit synapse-like characteristics—adjusting its state in response to external stimuli—and non-volatile properties to maintain stored data³⁶. To further ensure high-frequency processing capabilities and long-term system stability, we have designed floating gate synaptic transistors based on a two-dimensional van der Waals heterostructure as neuromorphic devices that generate temporal motion cues directly in hardware. Based on the Fowler-Nordheim tunneling mechanism, the floating gate synaptic transistor precisely regulates the charge in the floating gate through the gate voltage, thus realizing the long-term stable storage of information and guaranteeing the continuity and reliability of produced motion cues in the time dimension. Fig. 2a demonstrates a schematic structure of this floating gate synaptic transistor. From bottom to top, the synaptic transistor includes a gold film (serving as the control gate), an aluminium oxide (Al_2O_3) blocking layer, a multilayer graphene (MLG) floating gate, a thin hexagonal boron nitride (h-BN) tunneling layer, and a molybdenum disulfide (MoS_2) channel. In operation, gate-source voltage (V_{gs}) pulses are applied to the control gate (with the source grounded) to modulate the drain-source current (I_{ds}). Comprehensive details on the fabrication process of the floating gate synaptic transistor and the Raman characterization of its heterostructure are provided in Supplementary Figs. 1 and 2. Under these conditions, the MoS_2 channel's output characteristic confirms a good ohmic contact with the Cr/Au interface (Supplementary Fig. 3). In terms of memory behavior, the transfer curve of this synaptic transistor at a fixed drain-source voltage ($V_{\text{ds}} = 1\text{V}$) depicts a clockwise memory window that reaches 11.2 V when the V_{gs} is swept from -10 V to $+10$ V (Fig. 2b). Furthermore, the memory window increases with the maximum applied V_{gs} , as presented in Supplementary Fig. 4. When applying V_{gs} pulses, this synaptic transistor displays obvious synapse-like characteristics. As shown in Fig. 2c, the change in conductance is positively related to the number of applied pulses. The calculation method for the pulse-number linearity is detailed in the Supplementary Fig. 22. The modulation mechanism is elucidated by the energy band diagram: negative V_{gs} pulses drive holes into the floating gate and elevate the device's conductance, while positive V_{gs} pulses facilitate electrons tunneling into the floating gate and reduce its conductance. Additional details about the operating mechanism can be found in Supplementary Fig. 5. This modulation can be controlled by varying the amplitude and duration of applied voltage stimuli. As depicted in Fig. 2d, the increase in conductance correlates positively with both the amplitude and duration of the negative V_{gs} pulses. We successfully achieved weight regulation at a lower negative V_{gs} amplitude of 7 V by enhancing the gate coupling ratio^{37,38}, as detailed in Supplementary Note 7. Supplementary Fig. 6 illustrates the variation of conductance with the amplitude of the positive V_{gs} pulses. With respect to response speed, this floating gate

synaptic transistor demonstrates rapid operating speed, achieving a current switching of $60\ \mu\text{A}$ (from a low- to high-conductance state) under -15V V_{gs} pulse with a duration of $100\ \mu\text{s}$ (Fig. 2e). This $\sim 100\ \mu\text{s}$ response time is suitable for high-frequency visual information processing. Moreover, the synaptic transistor exhibits repeatable programming characteristics and multiple analog states (Fig. 2f), enabling the precise encoding of external information as its state. In terms of endurance, up to 8000 programming/erasing cycles can be achieved under positive and negative V_{gs} pulses ($\pm 15\text{V}$, 1 ms), with the I_{ds} at $V_{\text{ds}} = 1\text{V}$ remaining at $\sim 10^{-9}$ A and 10^{-5} A in the low- and high-conductance states, respectively (Fig. 2g). Furthermore, this synaptic transistor displays non-volatile behavior. Both the low- and high-conductance states are maintained for over 10^4 s (Fig. 2h), confirming their non-volatility in storing external stimuli data. The excellent retention characteristics are primarily attributed to the intrinsic physical properties of the material and the optimized device structure design, as detailed in the Supplementary Note 6. Compared with other reported devices (Fig. 2i), this synaptic transistor requires a low V_{gs} amplitude for weight modulation and presents excellent retention^{37,39–47}. The excellent performance of the device, including fast response, long endurance, and stable retention, can be attributed to the selection of materials and thicknesses of each functional layer, the enhanced gate coupling ratio³⁷, and the establishment of atomically sharp and flat interfaces^{39,48–50}; a detailed comparison is provided in Supplementary Fig. 7 and Supplementary Table 1. When scaling a single synaptic transistor to a 4×4 array as shown in Fig. 2j, the developed fabrication process can be found in the Methods section and Supplementary Fig. 8. After encapsulation (Fig. 2k), the array can be interfaced with external circuits via pins or connectors, facilitating integration with other system components. Such scalability paves the way for the development of commercial chips. The variation among multiple devices, as exhibited in Fig. 2l, demonstrates the consistent synaptic modulation behavior.

Temporal motion cue generated by neuromorphic devices

To directly generate temporal motion cues at the hardware level, we propose the imaging architecture illustrated in Fig. 3a. In this design, a conventional imaging array could serve as the front-end sensor, converting external stimuli into an analog voltage signal. This signal is then processed along two parallel paths: one is digitized to form the conventional image representation, while the other converts the analog signal and modulates a synapse array to record temporal information. Specifically, the circuit needed for voltage conversion, shown in Fig. 3b, c consists of a differential processing part and an amplitude conversion part. The differential processing part extracts changes in light intensity, while the amplitude conversion part generates voltage pulses applied to the synaptic transistor array, reflecting the temporal information of the current visual scene. In the differential processing part, a high-pass filter first differentiates the light intensity, and an operational amplifier (op-amp) then amplifies these changes within a suitable operating range for subsequent processing (Fig. 3b and Eq. 1):

$$V_{i,j}(t) = a \frac{dI_{i,j}}{dt} = a (I_{i,j}(t) - I_{i,j}(t - \Delta t)) \quad (1)$$

where V is the output voltage of the differential processing unit; Δt refers to the sampling interval; (i, j) represents the spatial coordinates of the temporal information, corresponding in size to the synapse array; I is the analog visual voltage transmitted from the imaging array (which may be resized to match the synapse array size); and a is a proportionality coefficient ensuring the voltage remains within a suitable operating range. In the following amplitude conversion part, an absolute circuit is constructed to extract the absolute voltage change, focusing on the magnitude of the light intensity change rather than its

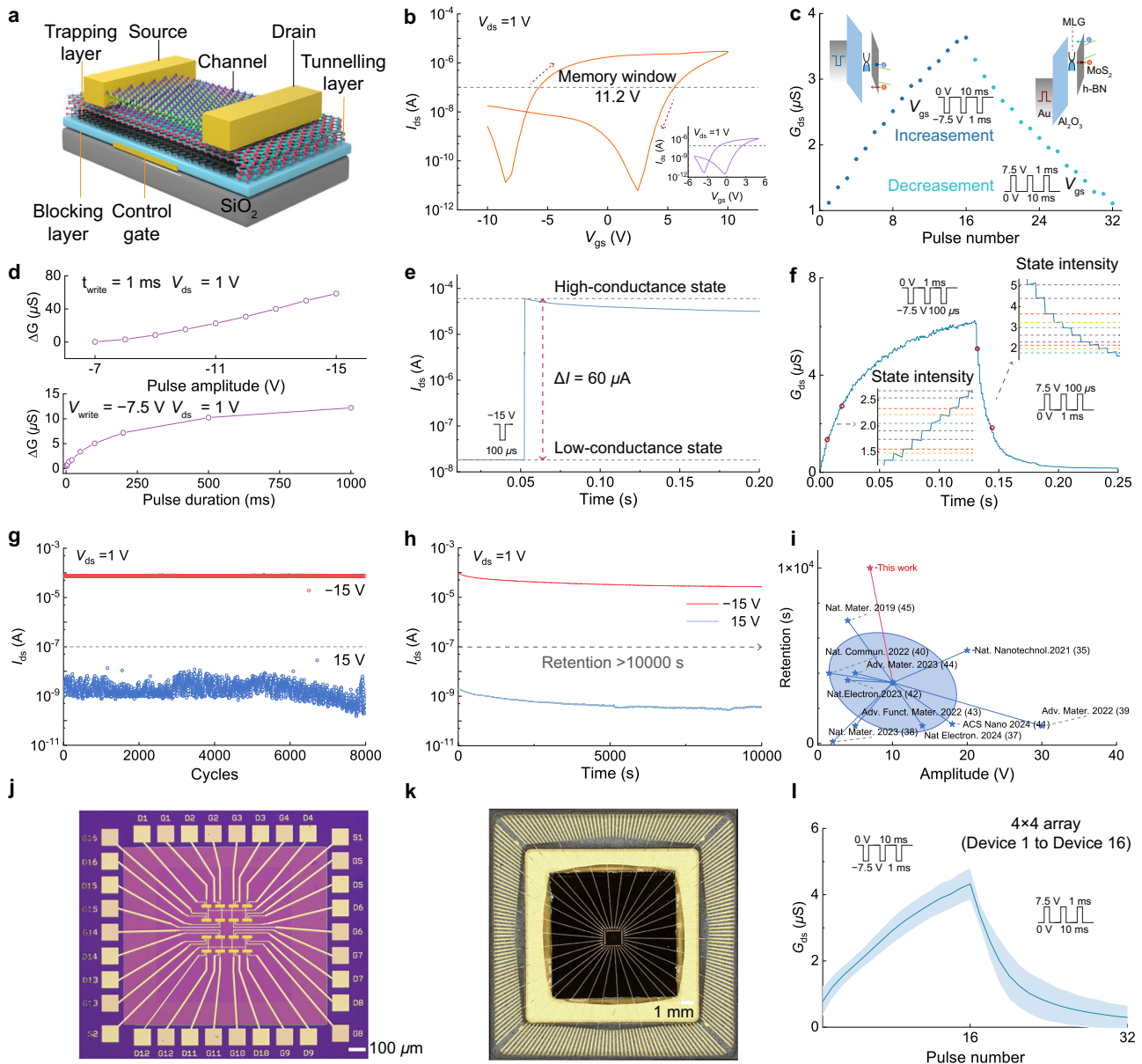


Fig. 2 | Performance characterization of the floating gate synaptic transistor. **a** Schematic illustration of the floating gate synaptic transistor structure. **b** Transfer characteristic of the device, with a memory window reaching 11.2 V at $V_{ds} = 1$ V when V_{gs} is swept from -10 V to 10 V. The inset shows the transfer curve for V_{gs} swept from -5 V to 5 V. **c** Conductance modulation through 16 negative/positive V_{gs} pulses with the operating mechanism of this synaptic transistor. **d** Absolute change in conductance (ΔG) as a function of V_{gs} pulse amplitude (top; pulse width = 1 ms) and pulse width (bottom, $V_{gs} = -7.5$ V). **e** Switching of conductance state under -15 V V_{gs} pulse of $100 \mu s$. **f** Conductance changes induced by 100 consecutive negative and positive V_{gs} pulses at 7.5 V amplitude and $100 \mu s$ width. The inset depicts multiple conductance states. **g** Endurance performance of the synaptic transistor executed with alternate positive and negative V_{gs} pulses (± 15 V, 1 ms). **h** Retention stability of

the synaptic transistor after 10 negative/positive V_{gs} pulse, demonstrating stable high- and low-conductance states. **i** Comparison with previously reported studies, where the range of the blue ellipse is derived from the mean and standard deviation of the data points in the most advanced research progress in recent years. The centroid represents the mean of the data, while the boundary of the ellipse corresponds to the 95% confidence interval. Further details can be found in Supplementary Table 1. **j, k** Optical images of the floating gate synaptic transistor array (**j**; scale bar, $100 \mu m$) and its bonded chip (**k**; scale bar, 1 mm). **l** Variations in pulse modulation among devices in the 4×4 array. Each device is modulated by 16 consecutive negative/positive V_{gs} pulses with the duration of 1 ms and amplitude of 7.5 V. The blue solid line and blue envelope represent the mean and variance of conductance of 16 devices under V_{gs} pulses.

direction (Eq. 2):

$$\hat{V}_{i,j} = |V_{i,j}| \quad (2)$$

Then, a reconfigurable op-amp generates a corresponding modulation pulse based on the current amplitude of the absolute voltage

$\hat{V}_{i,j}$. The relationship follows Eq. 3:

$$\tilde{V}_{i,j} = \begin{cases} p_1 \times (\hat{V}_{i,j} - b_1), \hat{V}_{i,j} > V_{th} \\ p_2 \times (\hat{V}_{i,j} - b_2), \hat{V}_{i,j} \leq V_{th} \end{cases} \quad (3)$$

where $\tilde{V}_{i,j}$ is the modulation voltage V_{gs} applied to the synapse array, p_1, p_2 are different proportional coefficients, b_1, b_2 represent different

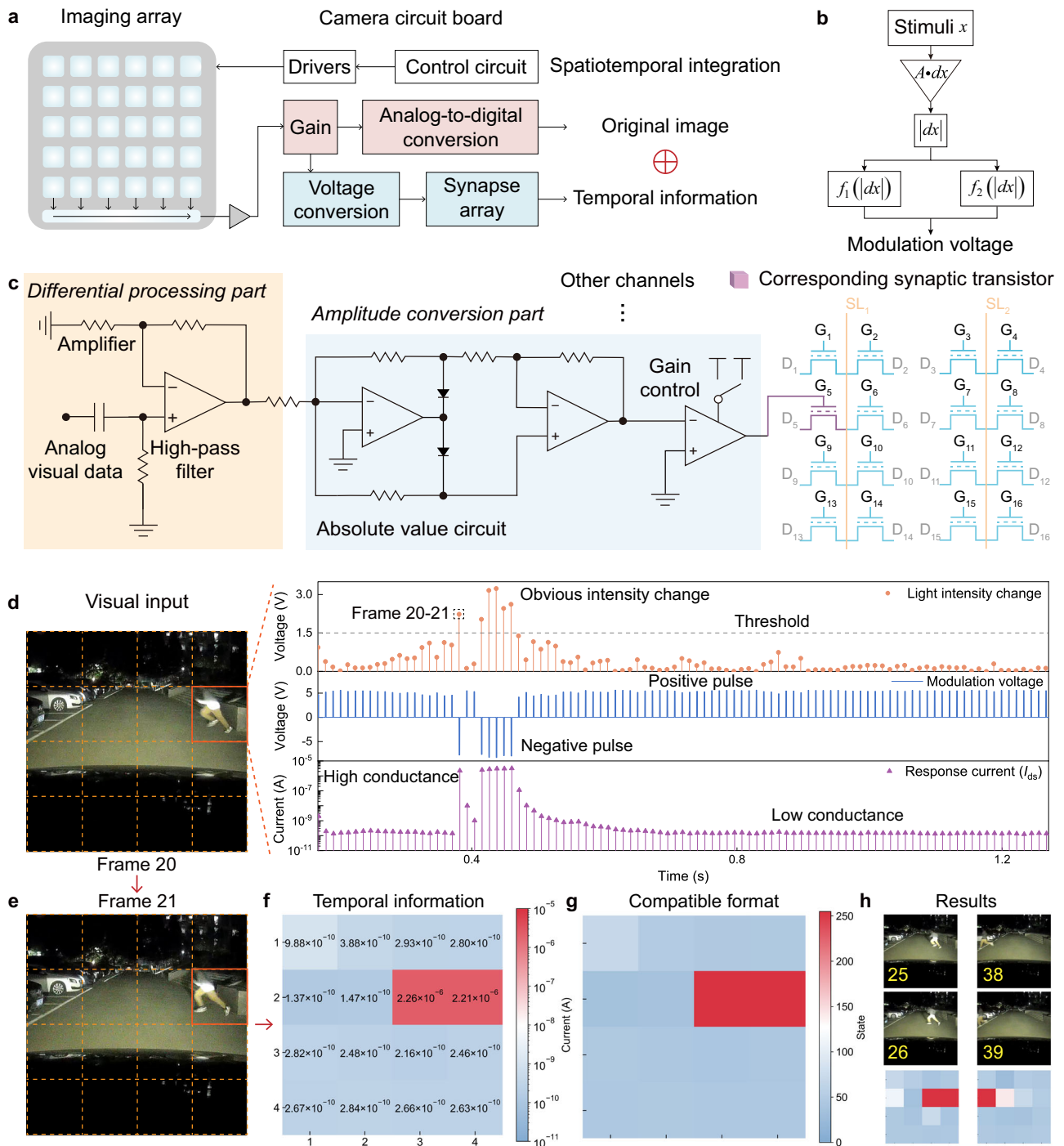


Fig. 3 | Imaging system architecture and temporal information extraction. **a** Architecture of the imaging system integrating temporal information with the original image. **b** Mathematical representation of the circuit used to extract temporal information. **c** Circuit diagram. **d** Visual input containing analog visual voltage changes and the corresponding V_{gs} applied to the synaptic transistor at location

(2, 4). **e** Visual input for the subsequent frame following (d). **f** Temporal information derived from the I_{ds} values of the synaptic transistor array states. **g** Temporal information scaled to the range of 0 to 255. **h** Examples of processed visual input in the vehicle scenario between frames 25 and 26, and between frames 38 and 39.

modulation biases, and V_{th} is a preset threshold. Under the effect of modulation voltage V_{gs} , the synaptic transistor at position (i, j) of array is modulated to a resistance state related to the nature of the light intensity change. By comparing it with the threshold V_{th} , dramatic changes caused by potential moving objects and mild changes caused by background movement or noise are separated and translated into negative and positive V_{gs} pulses, respectively, resulting in different trends of device state switching. The pulse width and the maximum

supported processing frequency are detailed in Supplementary Note 13. When analyzing this synapse array, the temporal motion information can be inferred from the distribution of I_{ds} values. For example, devices with high I_{ds} values (low-resistance under the modulation of negative V_{gs} pulses) that cluster spatially indicate regions containing moving objects. More details of this circuit and the manipulation of analog visual voltage can be found in Supplementary Fig. 10.

In our implementation, we employed a commercial camera to capture 800×800 images within a vehicle scenario and processed the visual input using the 4×4 synapse array (Fig. 3d). During processing, the visual stimuli captured by the commercial image sensor are resized by averaging the light intensity of a matrix of $m \times n$ pixels into a basic unit. Here, m and n are set to 200. In this configuration, external visual stimuli are translated into modulation voltages for the synapse array according to the voltage conversion circuit. Thus, the temporal information of the current visual scene is mapped onto this transistor array. For instance, when a pedestrian suddenly runs in front of the car (Fig. 3d, e), a noticeable change in the analog visual voltage occurs at position (2, 4), leading to a negative V_{gs} pulse. As a result, the synaptic transistor at (2, 4) switches to a high-conductance state, causing its I_{ds} value to increase significantly under a fixed V_{ds} . As shown in Fig. 3f, the distribution of measured I_{ds} across the 4×4 synaptic transistor array clearly reflects the temporal dynamics of the current visual scene, highlighting the presence of a moving object on the right side. To further facilitate integration with conventional visual processing methods, these temporal data are transformed into values ranging from 0 to 255, as shown in Fig. 3g using a logarithmic mapping (see “Methods”), making them compatible with common image processing libraries such as Python’s OpenCV (cv2) package. Through this pipeline, the temporal information encoded by the neuromorphic devices can be seamlessly combined with the original image (Fig. 3h).

Accelerated movement velocity calculation

After extracting temporal information of the current visual scene from the synapse array states, this data can be transformed into temporal cues that accelerate velocity estimation, ultimately producing the optical flow assisted by temporal cues (Fig. 4a). Specifically, the conversion into temporal cues involves two main steps. First, there is a binarization process based on a predefined threshold. Second, a connectivity analysis that includes defining connectivity, labeling connected components, and expanding these regions is performed (Fig. 4b). The resulting list of connected regions serves as the temporal cues, as shown in Fig. 4c. In addition to the vehicle operation scenario, the UAV operation scenario with a resolution of 160×160 is also processed using the same 4×4 synaptic transistor array. By comparing the original image and the generated temporal cues, it is observed that the constructed temporal cues areas effectively highlight potential moving objects. During the subsequent velocity inference, the temporal cues serve as ROIs that help automatically filter the areas where movement velocities need to be calculated (Fig. 4d). This filtering process speeds up velocity calculations compared to processing the entire image. Additionally, the neuromorphic pipeline seamlessly integrates with current velocity inference approaches, whether they are based on traditional computer vision techniques or neural networks.

In our implementation, we demonstrate the adaptability of our method using three representative algorithms for movement velocity calculation: the traditional Farneback algorithm and the neural network-based GMFlow and RAFT algorithms, as shown in Fig. 4e. Integration with other algorithms, such as FlowFormer, is demonstrated in Supplementary Figs. 11 and 13. These algorithms vary in their operational characteristics and should be selected based on the practical working environment. For instance, Farneback is suitable for less demanding scenarios, while GMFlow and RAFT are more appropriate for situations requiring higher accuracy and adaptability, albeit at increased computational and technical costs. Nevertheless, the capability of integrating various movement velocity calculation methods ensures that the neuromorphic pipeline is applicable to unstructured environments, which can vary significantly. Taking the vehicle operation scenario as an example, the temporal cues filter the visual input so that only highlighted regions are forwarded to the subsequent velocity inference stage. In practical implementations, to

improve robustness and resolution, the selected region is slightly padded to include peripheral information about the moving area. As shown in Fig. 4f, the detected motion region is slightly padded, and the movement information can be calculated using various velocity inference algorithms. Details on the strategy of padding and handling multiple moving objects can be found in Supplementary Note 5.

In the vehicle scenario, processing examples are shown in Fig. 4g. Even during periods of high motion—when the running pedestrian occupies a large portion of the scene—our method remains faster (Fig. 4i). Compared to conventional optical flow methods, our spatiotemporal approach enables the detection of potential motion regions (1–2 ms) as shown in Supplementary Fig. 15, and the average total velocity inference times for Farneback, GMFlow, and RAFT are reduced to 13.0%, 37.2%, and 19.6%, respectively (Fig. 4j). For the UAV scenario analysed with Farneback as shown in Fig. 4h, the average inference time is reduced to 51.0% of the original duration (Supplementary Fig. 12).

Demonstration of fundamental tasks

Optical flow assisted by temporal cues integrates both temporal and spatial motion information (e.g., calculated spatial movement velocity), and it can support fundamental tasks that enable autonomous vehicles, UAVs, and robots to perceive, understand, and interact with their environments autonomously and intelligently (Fig. 5a). During task execution, these motion cues are first decomposed. Then, temporal motion cues selectively filter spatial cues to focus only on regions with potential motion information rather than processing the entire scene. This selective focus significantly accelerates subsequent task execution. The filtered spatial cues are then combined with visual input to execute task-specific algorithms. The implementation pipeline is detailed in Supplementary Note 4. Aligning with the vision of utilizing optical flow to help robotic systems efficiently perceive the dynamic world, similar to biological visual systems, the above pipeline for processing visual information in our method mirrors the human visual system. This includes a perception unit corresponding to the human eye, a synapse array that extracts temporal motion cues analogous to the LGN, and task-specific algorithms that perform high-level processing similar to the visual cortex. Additional descriptions of the human visual system can be found in Supplementary Note 1.

In our implementation, the visual inputs encompass various scenarios, including vehicle operation, UAV operation, sports activities (e.g., table tennis from the UCF dataset)⁵¹, and grasp operation (Supplementary Movie 2). The settings for visual processing are provided in the Methods section and in Supplementary Table 5. Utilizing optical flow assisted by temporal cues, essential tasks such as motion prediction, object segmentation, and object tracking are performed on these visual inputs. Detailed task-specific algorithms can be found in the Methods section and Supplementary Fig. 16. As shown in Fig. 5b, spatial motion cues are calculated using multiple velocity inference methods and employed to execute various tasks. Additional results are presented in Supplementary Figs. 17–21. Notably, the table-tennis sequences did not exhibit missed detections despite the small object size, owing to an appropriately chosen sensor-to-synapse pooling size, as detailed in Supplementary Notes 11–12. The overall accelerated processing, which includes both velocity inference and task execution in the neuromorphic pipeline, achieves processing times comparable to human perception—approximately 150 ms (Supplementary Note 1).

In addition to high processing efficiency, these tasks are evaluated using standard metrics, including the Structural Similarity Index Measure (SSIM), pixel accuracy (PA), and Intersection over Union (IoU), and neuromorphic pipeline achieves comparable performance to conventional pipeline (Fig. 6a). In certain scenarios, such as vehicle operation, UAV operation (small), and grasping, neuromorphic pipeline significantly outperforms conventional methods. On average, the accuracy improvements are 213.5%, 157.4% and 740.9%, respectively.

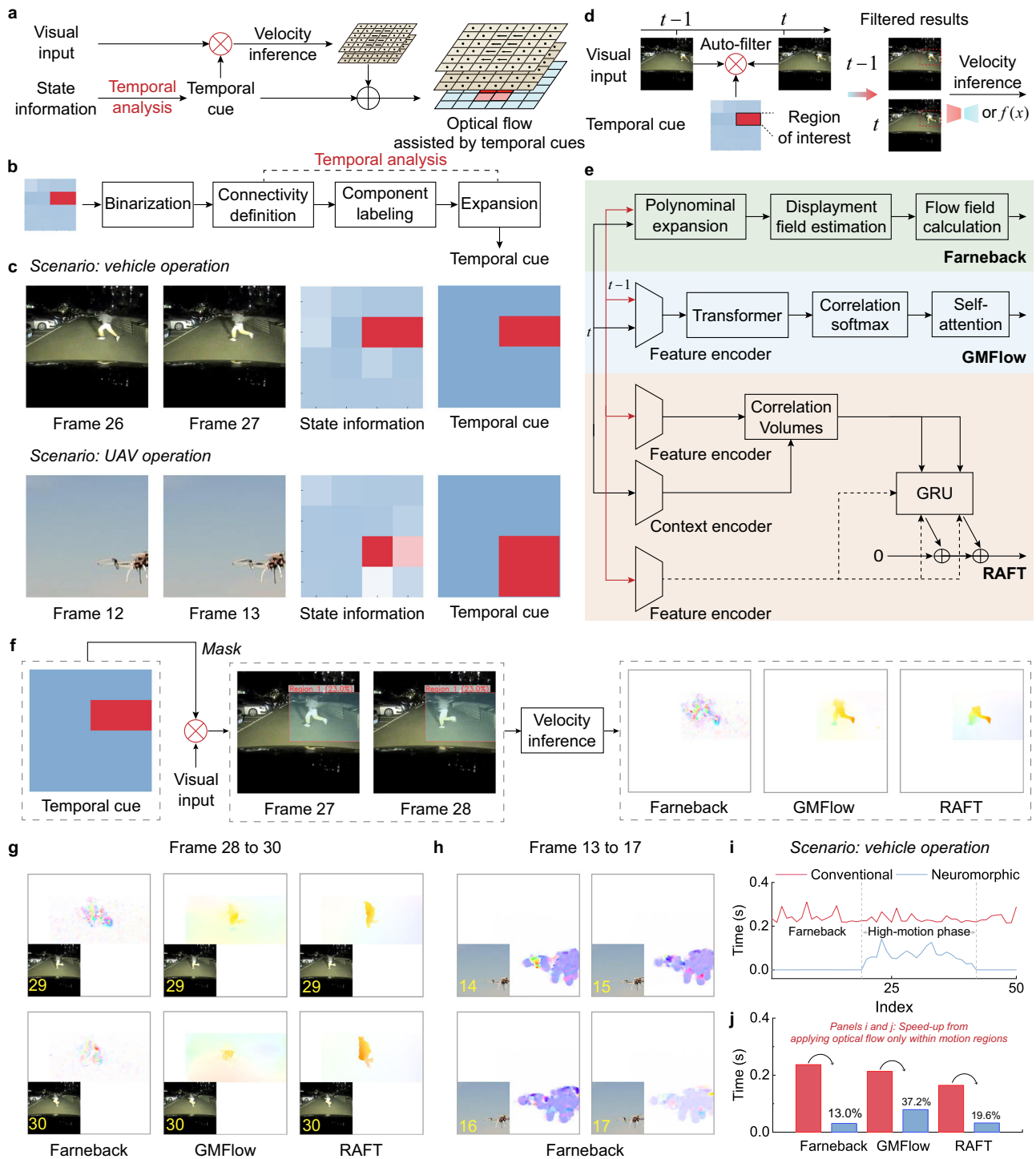


Fig. 4 | Optical flow generation assisted by temporal cues. **a** Algorithm structure for generating optical flow assisted by temporal cues. **b** Process of translating temporal information based on synapse array states into temporal motion cues. **c** Calculated temporal motion cues for vehicle and UAV operation scenarios. **d** Accelerated velocity inference using temporal cues. **e** Schematic diagram of different velocity inference algorithms, including Farneback, GMFlow, and RAFT. **f** Filtered visual input using temporal cues and resulted movement velocity. **g** Example results from the vehicle operation scenario. **h** Example results from the UAV operation scenario. **i** Comparison of velocity inference times between conventional and neuromorphic pipeline (using Farneback). **j** Average time comparison.

The comprehensive statistical metrics can be found in Supplementary Figs. 50–52, clearly indicating that the main source of the observed performance enhancement is the improved accuracy in object tracking tasks. Specifically, these performance improvements are attributed to the additional environmental knowledge embedded in the temporal cue. For instance, in the RAFT-based object segmentation for grasping

operations (Fig. 5b), RAFT cannot infer velocity accurately due to its limited generalization. However, the temporal cue provides a boundary constraint, enhancing segmentation accuracy. Similarly, in object tracking tasks, the temporal cue highlights the region containing the moving object while excluding irrelevant regions. As a result, the tracking accuracy is improved by reducing the impact of noise. The

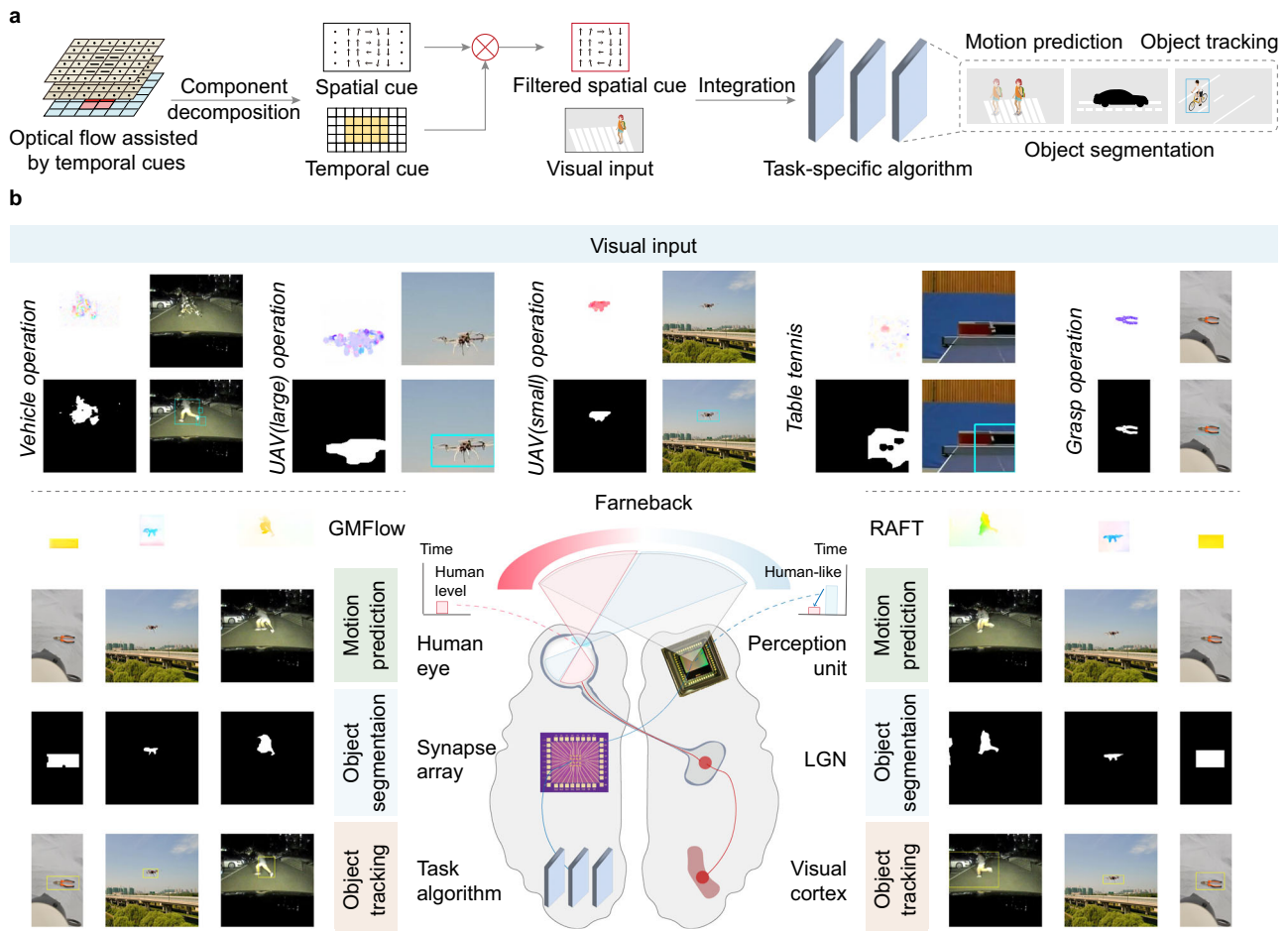


Fig. 5 | Demonstration of fundamental visual tasks. **a** Supported fundamental visual tasks. **b** Processing results of multiple scenarios including vehicle operation, UAV operation, sports activities, and grasp operations. The perception unit includes a CMOS sensor image by Filya1 (Wikimedia Commons), used under CC BY-SA 3.0.

results of significance testing, summarized in Supplementary Table 11, statistically confirm these improvements and underscore the robustness and reproducibility of our approach. More detailed task performance data can be found in Supplementary Table 2. Although device metrics such as uniformity and linearity (see Supplementary Table 6) are within commonly reported ranges for synaptic transistors, the demonstrated analog weight updating under cumulative same-polarity voltage pulses enables reliable storage of temporal motion patterns sufficient for our tasks. Design optimizations and algorithmic corrections used in the experiments to achieve such high accuracy can be found in Supplementary Note 8. In terms of processing efficiency (Fig. 6b), the ability of our method accelerates full-spectrum visual processing, including both velocity inference and task-specific algorithms. Thus, a faster response can be observed compared to conventional optical flow methods. When using Farneback for velocity inference, the acceleration ratio—i.e., the ratio of the original whole processing time to the whole accelerated processing time—ranges from 12.5 to 58.0%, with an average of 27.5%. For GMFlow-based tasks, the acceleration ratio ranges from 4.7 to 36.7%, with an average of 20.6%. For RAFT-based tasks, the acceleration ratio ranges from 16.7 to 53.3%, with an average of 29.1% (Supplementary Table 3). After examining all 33 groups of tasks, the average acceleration ratio is 26.1%, which corresponds to an approximate 4X speedup. When fitting the acceleration ratio of velocity inference and the percentage of filtered regions based on temporal cues in the vehicle scenario, clear linear relationships are observed for both Farneback and RAFT methods, with all R^2 values exceeding 0.94 (Fig. 6c). However, velocity inference using GMFlow does not follow this trend due to its unique operational

characteristics (Supplementary Fig. 14). Other velocity inference methods, such as FlowFormer, exhibit similar linear acceleration trends (Supplementary Fig. 13). In the acceleration of task execution, similar linear relationships can be observed using data points from Farneback-based vehicle scenario processing. This general acceleration, which correlates with the size of filtered visual input based on temporal cues, demonstrates the effectiveness of the proposed approach in enhancing both velocity inference and task execution performance. As a result, the neuromorphic motion extraction hardware pipeline enables real-time visual processing capabilities that are comparable to, or even exceed, human-level perception (Fig. 6d and Supplementary Table 4). A comparison of our method with other state-of-the-art neuromorphic visual approaches is provided in Supplementary Table 12.

Nevertheless, in challenging scenarios involving ego-motion or out-of-distribution motion, the performance of our method may degrade. To evaluate its robustness under such conditions, we conducted a comprehensive evaluation detailed in Supplementary Note 10 and Supplementary Movie 3. Specifically, we performed two sets of experiments: (i) controlled sequences captured using a hand-held phone, with concurrent IMU recordings of device motion, and (ii) real-world recordings from an in-car dashcam. In the first experiment, the observed speedup decreased to 170%, while in the second, the acceleration ratio was 74.8%, corresponding to a reduction in speed-up from 400 to 134% (1/74.8%). Regarding accuracy, as shown in Supplementary Table 8, performance degrades in complex scenes compared with sparse-motion scenarios. This is consistent with the limitations of conventional optical flow methods, which also perform poorly under

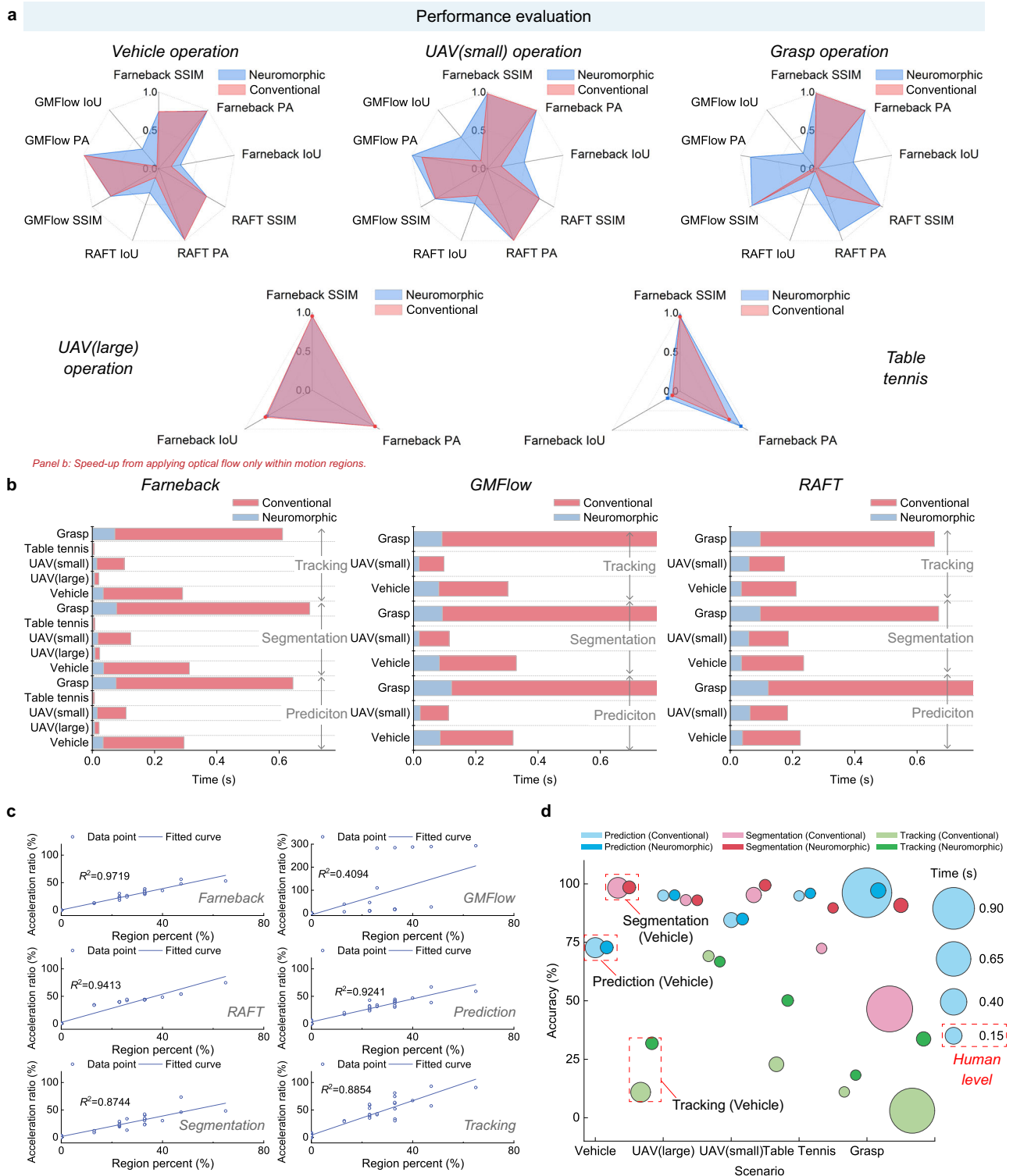


Fig. 6 | Performance evaluation and comparison. **a** Task performance comparison between conventional and neuromorphic pipeline. **b** Comparison of total processing time, including both velocity calculation and task execution, for conventional pipeline versus neuromorphic pipeline. **c** Acceleration characteristics of our method, illustrating the relationship between acceleration ratio and the

percentage of filtered regions based on temporal cues. **d** Comprehensive comparison of conventional and neuromorphic pipeline across various applications, including average task performance and processing time using different velocity inference methods (Farneback, GMFlow, and RAFT).

such conditions^{52,53}. The observed degradation arises primarily from the constraints of the deployed optical flow computation and downstream task algorithms. The core role of our neuromorphic hardware is to extract motion regions and generate ROIs, thereby accelerating

downstream processing, making it powerful in scenes with sparse motion. Furthermore, we discuss potential motion compensation and fallback strategies that could be incorporated to further enhance system robustness (Supplementary Note 10).

Discussion

Compared to conventional spatial-only optical flow methods, optical flow assisted by temporal cues integrates additional temporal motion cues of the current visual scene. By utilizing the spatial-temporal consistency of motion, which refers to the simultaneous spatial displacement of pixels and the temporal variation in light intensity within a motion region, this added temporal information enables the direct delineation of potential moving regions in as little as 1–2 ms using synapse array state information. This delineation offers two major benefits. First, it enables selective processing of visual input, resulting in substantially faster velocity calculations and task execution. Second, the delineation information provides valuable prior knowledge for velocity inference and task execution processes. For instance, in object tracking tasks, the temporal cues from our approach constrain the tracking range, reducing false detections from background noise and greatly enhancing robustness. Furthermore, for neural network-based velocity inference, the delineation information supplies a reasonable range of results even in untrained working environments, thus addressing the limited generalization problem of current neural network methods.

This ROI-first strategy has precedents in software and sensor work that restrict optical-flow computation to likely moving regions. For example, Sagar et al. detect foreground regions and combine foreground-focused processing with template/scale cues for monocular MAV obstacle avoidance; their pipeline demonstrates that processing only foreground regions can substantially reduce computation and improve obstacle confidence in constrained flight settings⁵⁴. Denman et al. propose a combined optical-flow algorithm that uses foreground masks to limit flow computations and to improve flow near object boundaries⁵⁵. Our work differs from these prior software pipelines in one fundamental respect: the temporal cues are generated directly in analog hardware by synaptic transistors with non-volatile, high-frequency response. This hardware generation obviates repeated frame-to-frame accumulation in software, enabling rapid ROI formation and thus ultrafast visual perception. In short, while Sagar & Visser and Denman et al. demonstrate the value of restricting flow to motion regions, our synaptic array provides a hardware temporal attention mechanism that produces those regions faster and with a different information representation.

The primary novelty of our work is the proposed framework, which couples temporal information generated in situ by the neuromorphic synapse array with spatial gradients extracted from image frames, thereby accelerating the entire pipeline—from velocity inference to downstream high-level processing tasks. This framework is general and can be integrated with other neuromorphic synaptic devices, including ferroelectric memristors, phase change memory, etc. Through the non-volatile characteristics of synaptic devices, our method enables efficient temporal cue generation directly in hardware, significantly reducing external latency compared to software-only approaches (see Supplementary Note 14). The overall pipeline latency is shown in Supplementary Fig. 49. Besides, this pipeline can also be realized with photo-memory devices; a comparison between our separated design and integrated photo-memory implementations is provided in Supplementary Note 9.

In practical implementation, the neuromorphic motion extraction hardware pipeline significantly reduces the processing time of visual data, enabling robotics to excel in more complex tasks, particularly those requiring real-time processing capabilities like collision avoidance and object tracking. For example, in vehicle operations, the average -0.2 s improvement in processing time observed in our method can result in a reduction in full-braking distance of 4.4 m at a speed of 80 km h⁻¹, greatly enhancing driving safety. Similarly, our method enables at least a threefold reduction in reaction time in UAV (small) scenarios, significantly improving their durability and performance in dynamic environments. Across all tasks using

Farneback and GMFlow for velocity inference, processing times remain below 40 ms. This enhanced processing allows UAVs to track moving objects between frames in the setting of 25 frames per second. As a result, UAVs can adjust their speed and pose in real-time, achieving near-theoretical minimum delay in target tracking. Beyond robotic applications, our method holds great promise for improving human-robot interaction. With an emphasis on response time to ensure real-time feedback, robots must interpret visual scenes—such as gestures and movement recognition—within 100 to 200 ms. The ultrafast visual processing enabled by neuromorphic motion extraction hardware can serve as a crucial information source for future human-robot interaction, ensuring smooth and responsive engagements.

Looking forward, the core principle of our approach lies in capturing the temporal information of visual scenes through synapse arrays, enabling a temporally guided analysis of visual stimuli. This design ensures high compatibility with various types of front-end sensors. Compared with event-based cameras, which also extract motion changes rapidly, the proposed synapse array performs analog accumulation of light-intensity changes, producing a continuous-valued representation that reflects short-term motion history. Thus, the synapse array should be viewed as an alternative approach to event-based vision for detecting regions of motion. Its output, determined by the continuous conductance states of the neuromorphic hardware, differs fundamentally from the binary ON/OFF events produced by DVS sensors, providing cleaner and more actionable data for high-level processing beyond mere optical flow computation. A more detailed comparison is provided in Supplementary Note 15, and we further demonstrate that our neuromorphic hardware can accumulate event-based input streams (see Supplementary Note 16). In terms of applications, this temporally guided approach extends far beyond optical flow calculation alone. For instance, after identifying potential ROI within our proposed system architecture, other algorithms—such as YOLO neural networks for object detection—can be directly applied to these identified areas (Supplementary Fig. 53). As a result, computational resource usage and the time required for visual motion processing are minimized. With the ability to enhance efficiency across a wide range of applications, our spatial-temporal integrated approach could pave the way for broader adoption in fields such as robotics, autonomous systems, and computer vision, driving transformative advancements.

In conclusion, this work proposes and demonstrates a neuromorphic motion extraction hardware pipeline leveraging a synapse array to deliver a more comprehensive and efficient understanding of visual scenes compared to conventional spatial-only optical flow approaches. Compared to conventional optical flow methods, our method encodes additional temporal motion cues directly within the hardware, which identify ROIs in real time. Therefore, the full spectrum of optical flow-based visual processing can be accelerated. Furthermore, the seamless integration of various movement velocity calculation algorithms ensures adaptability across real-world complex environments. Benchmark evaluations across multiple robotic platforms and tasks demonstrate that our method outperforms state-of-the-art algorithms, achieving an average 4X improvement in processing speed while maintaining or enhancing accuracy in motion prediction, object tracking, and segmentation. Notably, our method facilitates the entire processing time, including velocity inference and task execution that approach or exceed human-level speeds (approximately 150 ms), thereby realizing the initial vision of optical flow and providing autonomous systems with unparalleled perception capabilities essential for safe and intelligent interaction with dynamic environments.

Methods

Device fabrication

The bottom control gates in the floating gate synaptic transistors were prepared via electron beam lithography (EBL) on 285 nm SiO₂/Si substrate, followed by thermal evaporation. Atomic layer deposition (ALD) technique was used to deposit the gate dielectric aluminum oxide (Al₂O₃). MoS₂/h-BN/MLG heterostructure in a single device was prepared by mechanical exfoliation from bulk materials (Nanjing MKNANO Tech. Co., Ltd., <https://www.mukenano.com>), and precisely positioned on the Al₂O₃ via the dry-transfer method. MoS₂/h-BN/MLG structures in a floating gate synaptic transistor array were fabricated from chemical vapor deposited materials (Six Carbon Technology Shenzhen), and stacked via wet transfer, then patterned through reaction ion etching according to the structural design. Furthermore, 5/50 nm Cr/Au drain-source electrodes on MoS₂ channel were defined sequentially by EBL, thermal evaporation and lift-off.

Device characterization

Electrical measurements of the floating gate synaptic transistor were conducted by a semiconductor parameter analyzer (B1500A, Keysight) under atmospheric conditions. The thickness of the two-dimensional material was measured by Bruker Mulyi-Mode 8 AFM. Raman characterization was tested using a confocal Raman spectrometer (WITec alpha 300 R) with a 532 nm laser as the excitation source.

Logarithmic mapping

To transform the drain-source current of the floating gate transistor to a range of 0–255, a logarithmic mapping is employed as shown in Eq. 4:

$$s = -\frac{3366}{\log_{10}^{I_{ds}}} - 306 \quad (4)$$

where I_{ds} represents the drain-source current, and s is the transformed current scaled within the range of 0 to 255.

Task algorithms

In the task of motion prediction, it involves predicting the position of a moving object. For the filtered visual input, by using reference frames and employing the Lanczos interpolation method, the moving object in the next moment can be inferred. The Lanczos interpolation formula is as Eqs. 5 and 6:

$$g(x) = \sum_{k=-n}^n g(k)L_n(x-k) \quad (5)$$

$$L_n(x) = \begin{cases} \frac{\sin(\pi x) \sin(\frac{\pi x}{n})}{(\pi x)^2}, & x \neq 0 \\ 1, & x = 0 \end{cases} \quad (6)$$

where $g(x)$ represents the interpolated value at position x , $g(k)$ represents the pixel values at integer positions k in the current frame, $L_n(x)$ is the Lanczos kernel function, and n is the size of the kernel.

For object segmentation, the optical flow assisted by temporal cues is first converted to polar coordinates. This transformation separates the original movement velocity information into direction (angle) and magnitude (distance) components, making it easier to analyze motion information. Notably, this step only manipulates the ROI that includes significant moving objects inferred from the motion pattern layer, thereby omitting regions with slight noise caused by environmental changes, such as slow variations in lighting. After the transformation, the image is converted from RGB to HSV color space, where the direction and magnitude of motion velocity layers are represented using the hue and value channels, respectively. This process is beneficial for subsequent processing because it allows more intuitive manipulation of color-based information: the hue channel

encodes the direction of motion, while the value channel encodes the magnitude of motion. This separation simplifies the process of identifying and segmenting moving objects based on their motion characteristics. When the motion information of the ROI is represented in the HSV color space, thresholding operations along with erosion and dilation operations can be applied to create a binary mask that accurately segments the moving objects. Thresholding isolates the relevant motion information, while erosion and dilation help refine the segmentation by removing small noise and closing gaps in the detected objects, respectively. This process results in a clear and precise segmentation of moving objects, thus enabling subsequent tasks such as object tracking and interaction in dynamic environments.

In object tracking, coordinate conversion, which is similar to that used in object segmentation, is applied first. This conversion separates motion information into direction and magnitude components. Following this, morphological opening is performed to smooth boundaries and remove noise. Then, using the contour detection algorithm, multiple bounding boxes are detected. Next, non-maximum suppression is applied to eliminate redundant detections and retain the most significant objects. This step ensures that only the most prominent moving objects are tracked across frames, improving the accuracy of the tracking process. Unlike conventional optical flow, which can be disturbed by background movements leading to unnecessary tracking, our pipeline focuses solely on the ROI regions that include potential moving objects. This targeted approach enhances tracking precision and reduces computational overhead by ignoring irrelevant background motion.

When evaluating the performance of the above tasks, metrics including SSIM, PA, and IoU are calculated to quantify the quality of prediction, segmentation, and tracking (Eq. 7–10), respectively:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x and μ_y are the average values of predict result x and ground truth y , σ_x^2 and σ_y^2 are the variances of x and y , σ_{xy} is the covariance of x and y , and C_1 and C_2 are constants to stabilize the division with weak denominator.

$$PA = \frac{\sum_i n_{ii}}{\sum_i t_i}$$

where n_{ii} is the number of pixels correctly classified for class i , and t_i is the total number of pixels in class i . Here, PA calculates the percentage of correctly segmented pixels.

$$IoU = \frac{\sum_{i=1}^n IoU_{si}}{n}$$

$$IoU_s = \frac{|A \cap B|}{|A \cup B|}$$

where $|A \cap B|$ is the area of overlap between the tracking mask A and the ground truth mask B , and $|A \cup B|$ is the area of union between the tracking mask A and the ground truth mask B , IoU_{si} represents the IoU_s between the i -th bounding box and the ground truth. All evaluation metrics (SSIM, PA, and IoU) are calculated over the entire image area for both conventional and our methods to ensure fair comparison.

Running environment

Performance evaluations of the neuromorphic pipeline with the Farneback method for velocity inference are performed on the 12th Generation Intel® Core™ i9-12900H processor. In contrast, performance evaluations of the neuromorphic pipeline utilizing neural

network-based velocity inference methods, including RAFT, GMFlow and FlowFormer, are conducted on a server outfitted with an NVIDIA V100 GPU and an Intel® Xeon® Platinum 8260 CPU operating at 2.40 GHz.

Visual processing

To demonstrate the scalability of our approach, visual input data—encompassing UAV (small) operations, table tennis, and grasping scenarios—are simulated using the synapse array based on our fabricated synaptic transistor (Supplementary Fig. 9).

Visualization of optical flow

Following the work by Baker et al., optical flow vectors are mapped to a color-coded image⁵². In this visualization approach, color hue represents the direction of motion, and color intensity/saturation corresponds to the magnitude of motion.

Data availability

All data supporting this study and its findings are available within the article, its Supplementary Information and associated files. Source data have been deposited in Figshare under accession code <https://doi.org/10.6084/m9.figshare.30977674>.

Code availability

All the necessary codes used in the tactile experiments and visual experiments, and their descriptions are available in <https://github.com/RTCartist/Neuromorphic-Spatiotemporal-Optical-Flow>.

References

- Gibson, J. J. *The Perception of the Visual World* p. 242 (Houghton Mifflin, 1950).
- Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.* **61**, 304–314 (1954).
- Gibson, J. J. Optical motions and transformations as stimuli for visual perception. *Psychol. Rev.* **64**, 288–295 (1957).
- Horn, B. K. P. & Schunck, B. G. Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981).
- Guizilini, V., Lee, K.-H., Ambruş, R. & Gaidon, A. Learning optical flow, depth, and scene flow without real-world labels. *IEEE Robot. Autom. Lett.* **7**, 3491–3498 (2022).
- de Croon, G. C. H. E., De Wagter, C. & Seidl, T. Enhancing optical-flow-based control by learning visual appearance cues for flying robots. *Nat. Mach. Intell.* **3**, 33–41 (2021).
- Teed, Z. & Deng, J. RAFT: Recurrent all-pairs field transforms for optical flow. in *Computer Vision—ECCV 2020* (eds Vedaldi, A., Bischof, H., Brox, T. & Frahm, J.-M.) 402–419. https://doi.org/10.1007/978-3-030-58536-5_24 (Springer International Publishing, 2020).
- Huang, Z. et al. FlowFormer: a transformer architecture for optical flow. in *Computer Vision—ECCV 2022* (eds Avidan, S., Brostow, G., Cissé, M., Farinella, G. M. & Hassner, T.) 668–685. https://doi.org/10.1007/978-3-031-19790-1_40 (Springer Nature, 2022).
- Xu, H., Zhang, J., Cai, J., Rezatofighi, H. & Tao, D. GMFlow: learning optical flow via global matching. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (eds Dana, K. et al.) 8121–8130 (IEEE, 2022).
- Grigorescu, S., Trasnea, B., Cocias, T. & Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot.* **37**, 362–386 (2020).
- Xu, H., Chen, J., Meng, S., Wang, Y. & Chau, L.-P. A survey on occupancy perception for autonomous driving: The information fusion perspective. *Inf. Fusion* **114**, 102671 (2025).
- Hagenaars, J., Paredes-Valles, F. & de Croon, G. Self-supervised learning of event-based optical flow with spiking neural networks. *Adv. Neural Inf. Process. Syst.* **34**, 7167–7179 (2021).
- Zhao, S., Zhao, L., Zhang, Z., Zhou, E. & Metaxas, D. Global matching with overlapping attention for optical flow estimation. In *Proc 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 17592–17601 (IEEE, 2022).
- Buades, A., Lisani, J.-L. & Miladinović, M. Patch-based video denoising with optical flow estimation. *IEEE Trans. Image Process.* **25**, 2573–2586 (2016).
- Liu, H., Hong, T.-H., Herman, M., Camus, T. & Chellappa, R. Accuracy vs efficiency trade-offs in optical flow algorithms. *Comput. Vis. Image Underst.* **72**, 271–286 (1998).
- Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am A.* **2**, 284–299 (1985).
- Borst, A. & Helmstaedter, M. Common circuit design in fly and mammalian motion vision. *Nat. Neurosci.* **18**, 1067–1076 (2015).
- O'Connor, D. H., Fukui, M. M., Pinsk, M. A. & Kastner, S. Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.* **5**, 1203–1209 (2002).
- McAlonan, K., Cavanaugh, J. & Wurtz, R. H. Guarding the gateway to cortex with attention in visual thalamus. *Nature* **456**, 391–394 (2008).
- Clifford, C. W. G. & Ibbotson, M. R. Fundamental mechanisms of visual motion detection: models, cells and functions. *Prog. Neurobiol.* **68**, 409–437 (2002).
- Zidan, M. A., Strachan, J. P. & Lu, W. D. The future of electronics based on memristor systems. *Nat. Electron* **1**, 22–29 (2018).
- Wang, S. et al. Memristor-based adaptive neuromorphic perception in unstructured environments. *Nat. Commun.* **15**, 4671 (2024).
- Yoon, J. H. et al. An artificial nociceptor based on a diffusive memristor. *Nat. Commun.* **9**, 417 (2018).
- Hong, S. J. et al. Bio-inspired artificial mechanoreceptors with built-in synaptic functions for intelligent tactile skin. *Nat. Mater.* 1–9 <https://doi.org/10.1038/s41563-025-02204-y> (2025).
- Donati, E. & Valle, G. Neuromorphic hardware for somatosensory neuroprostheses. *Nat. Commun.* **15**, 556 (2024).
- Truong, S. N., Ham, S.-J. & Min, K.-S. Neuromorphic crossbar circuit with nanoscale filamentary-switching binary memristors for speech recognition. *Nanoscale Res. Lett.* **9**, 629 (2014).
- Seo, S. et al. Artificial van der Waals hybrid synapse and its application to acoustic pattern recognition. *Nat. Commun.* **11**, 3936 (2020).
- Gao, S. et al. Programmable Linear RAM: a new flash memory-based memristor for artificial synapses and its application to a speech recognition system. In *Proc. 2019 IEEE International Electron Devices Meeting (IEDM)* 14.1.1–14.1.4. <https://doi.org/10.1109/IEDM19573.2019.8993598> (2019).
- Lee, J. et al. Light-enhanced molecular polarity enabling multi-spectral color-cognitive memristor for neuromorphic visual system. *Nat. Commun.* **14**, 5775 (2023).
- Choi, C. et al. Curved neuromorphic image sensor array using a MoS₂-organic heterostructure inspired by the human visual recognition system. *Nat. Commun.* **11**, 5934 (2020).
- Huang, H. et al. Fully integrated multi-mode optoelectronic memristor array for diversified in-sensor computing. *Nat. Nanotechnol.* **20**, 93–103 (2025).
- Baek, E. et al. Neuromorphic dendritic network computation with silent synapses for visual motion perception. *Nat. Electron* **7**, 454–465 (2024).
- Chen, W. et al. Essential characteristics of memristors for neuromorphic computing. *Adv. Electron. Mater.* **9**, 2200833 (2023).
- Wang, S. et al. Memristor-based intelligent human-like neural computing. *Adv. Electron. Mater.* **9**, 2200877 (2023).
- Liu, F. et al. Printed synaptic transistor-based electronic skin for robots to feel and learn. *Sci. Robot.* **7**, eab17286 (2022).
- Zhang, W. et al. Neuro-inspired computing chips. *Nat. Electron* **3**, 371–382 (2020).

37. Liu, L. et al. Ultrafast non-volatile flash memory based on van der Waals heterostructures. *Nat. Nanotechnol.* **16**, 874–881 (2021).
38. Yu, J. et al. Simultaneously ultrafast and robust two-dimensional flash memory devices based on phase-engineered edge contacts. *Nat. Commun.* **14**, 5662 (2023).
39. Jiang, Y. et al. A scalable integration process for ultrafast two-dimensional flash memory. *Nat. Electron* **7**, 868–875 (2024).
40. Kang, J.-H. et al. Monolithic 3D integration of 2D materials-based electronics towards ultimate edge computing solutions. *Nat. Mater.* **22**, 1470–1477 (2023).
41. Lai, H. et al. Photoinduced multi-bit nonvolatile memory based on a van der Waals heterostructure with a 2D-perovskite floating gate. *Adv. Mater.* **34**, 2110278 (2022).
42. Li, G. et al. Photo-induced non-volatile VO₂ phase transition for neuromorphic ultraviolet sensors. *Nat. Commun.* **13**, 1729 (2022).
43. Lu, H., Wang, Y., Han, X. & Liu, J. An ultrafast multibit memory based on the ReS₂/h-BN/Graphene heterostructure. *ACS Nano* **18**, 23403–23411 (2024).
44. Migliato Marega, G. et al. A large-scale integrated vector–matrix multiplication processor based on monolayer molybdenum disulfide memories. *Nat. Electron* **6**, 991–998 (2023).
45. Yang, Q. et al. Controlled optoelectronic response in van der Waals heterostructures for in-sensor computing. *Adv. Funct. Mater.* **32**, 202207290 (2022).
46. Zha, J. et al. A 2D heterostructure-based multifunctional floating gate memory device for multimodal reservoir computing. *Adv. Mater.* **36**, 2308502 (2024).
47. Zhu, X., Li, D., Liang, X. & Lu, W. D. Ionic modulation and ionic coupling effects in MoS₂ devices for neuromorphic computing. *Nat. Mater.* **18**, 141–148 (2019).
48. Wu, L. et al. Atomically sharp interface enabled ultrahigh-speed non-volatile memory devices. *Nat. Nanotechnol.* **16**, 882–887 (2021).
49. Huang, X. et al. An ultrafast bipolar flash memory for self-activated in-memory computing. *Nat. Nanotechnol.* **18**, 486–492 (2023).
50. Wang, H. et al. Ultrafast non-volatile floating-gate memory based on all-2D materials. *Adv. Mater.* **36**, 2311652 (2024).
51. Soomro, K., Zamir, A. R. & Shah, M. UCF101: a dataset of 101 human action classes from videos in the wild. Preprint at <https://doi.org/10.48550/arXiv.1212.0402> (2012).
52. Baker, S. et al. A database and evaluation methodology for optical flow. *Int. J. Comput. Vis.* **92**, 1–31 (2011).
53. Wang, Y. et al. Occlusion-aware unsupervised learning of optical flow. In *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 4884–4893. <https://doi.org/10.1109/CVPR.2018.00513> (IEEE, 2018).
54. Sagar, J. & Visser, A. Obstacle avoidance by combining background subtraction, optical flow and proximity estimation. In *Proc. Int. Micro Air Vehicle Conf. and Competition (IMAV 2014)*. (Delft University of Technology, 2014).
55. Denman, S., Fookes, C. & Sridharan, S. Improved simultaneous computation of motion detection and optical flow for object tracking. In *Proc. 2009 Digital Image Computing: Techniques and Applications*. 175–182. <https://doi.org/10.1109/DICTA.2009.35> (2009).

Acknowledgements

S.G. and L.T. acknowledge support from the National Key Research and Development Program of China (2023YFB3208003 and

2023YFB3208002). L.T. also acknowledges support from the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China, the Analysis & Testing Center and the start-up fund at the Beijing Institute of Technology. R.D. received no specific funding for this work, and the research in the paper is based on unfunded collaboration solely for the manuscript.

Author contributions

S.W., J.Z., T.P. and L.Z. contributed equally to the work. S.G. and S.W. conceived the idea and proposed the research. S.W. and T.P. designed the neuromorphic pipeline, with T.P. developing task-specific algorithms. L.Z. and S.W. evaluated the neuromorphic approach across various scenarios. X.G. collected the datasets and conducted data pre-processing. L.T. conceived the floating gate synaptic transistor design. J.Z., Y.C. and L.T. fabricated, tested, and analyzed the synaptic transistors. S.G., L.T. and X.G. supervised the project. S.G., S.W., L.T. and J.Z. wrote the manuscript with inputs from all authors. All authors discussed the results and implications and commented on the manuscript at all stages.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-68659-y>.

Correspondence and requests for materials should be addressed to Xiaoyu Guo, Li Tao or Shuo Gao.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026