Classification of Human Activity by Spiking Neural Networks using Event-based Vision Sensors

Aryan Rakheja
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Rohan Saini
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Ryuta Toyoda
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yuichiro Tanaka
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: rakheja-aryan859@mail.kyutech.jp, saini-rohan413@mail.kyutech.jp, toyoda.ryuta785@mail.kyutech.jp,
tanaka-yuichiro@brain.kyutech.ac.jp, tamukoh@brain.kyutech.ac.jp

Abstract

This paper presents a human action classification system using a spiking neural network (SNN) and an event-based vision sensor (EVS). The EVS captures asynchronous data with high temporal resolution, wide dynamic range, and motion blur resistance. SNNs, inspired by biological neurons, process this data event-driven, ensuring energy efficiency, low latency, and scalability. Using a custom EVS dataset of 600 videos across four action types and the optical flow for feature extraction, the system achieved 93% accuracy, offering an efficient solution for action recognition.

Keywords: Spiking neural network (SNN), Event-based vision sensor (EVS)

1. Introduction

Human action classification involves automatically identifying actions like running, walking, or gesturing from visual input, with applications in surveillance, healthcare, and human-computer interaction. Traditional methods often face challenges like motion blur, limited dynamic range, and high computational demands, especially in dynamic or low-light environments. Event-based vision sensors (EVS) provide solution by capturing asynchronous data triggered by scene changes, offering high temporal resolution, a wide dynamic range, and resistance to motion blur. These features make EVS ideal for human action classification [1]. In this paper, we propose integrating an EVS with a spiking neural network (SNN), a biologically inspired method that processes sparse, event-driven data, reducing computational complexity and energy consumption [2].

Recent studies have explored integrating EVS with SNNs for tasks like object and action recognition. One approach uses a hierarchical SNN architecture for action recognition, showing effectiveness on the dynamic vision sensor dataset. The SLAYER (spike layer error reassignment) training mechanism has achieved high gesture recognition accuracy, while another study introduces a spiking convolutional recurrent neural network to capture spatial and temporal correlations in event-based data [3], [4]. Our work uses the EVS with a feature extraction by the optical flow for efficient human action classification.

2. Methodology

2.1. *Data collection using event-based vision sensor*

We collected the dataset for human action classification using the EVS, capturing asynchronous event data from dynamic scenes. It comprises four actions: running, jogging, walking, and boxing, recorded under varied lighting and motion conditions to replicate real-world environments. To ensure diverse perspectives for each action, we positioned the sensor at specific angles. The blue parts indicate negative events that fall below the negative event threshold, while the red parts represent positive events exceeding the positive threshold, as shown in Fig. 1.

2.2. *Feature Extraction and Preprocessing*

We used the optical flow to extract motion-related features from the asynchronous event data. The optical flow measures the motion of objects between consecutive events by analyzing changes in pixel intensity over time.

The equation for estimating optical flow between two consecutive event frames is given as follows,

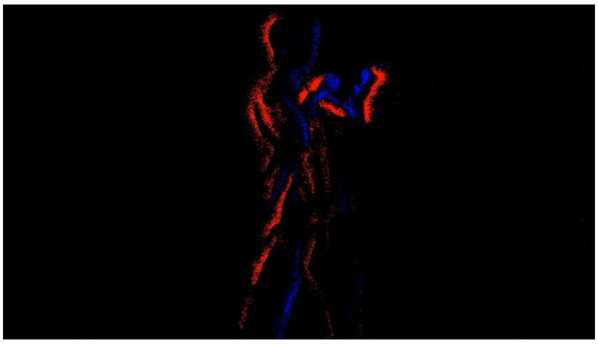


Fig. 1. Boxing action recorded by EVS.

$$I_t + I_x u + I_y v = 0 \tag{1}$$

where the temporal derivative of image intensity is denoted as I_t. I_x and I_y are the spatial derivatives, and u and v represent the velocity components in the x and y directions, respectively.

2.3. *SNN for Event-Driven Classification*

We used an SNN to process the event-driven data. In SNNs, neurons accumulate input over time and fire when the membrane potential exceeds a threshold. The membrane potential of the i-th neuron at time t is updated as:

$$V_i(t) = V_i(t-1) + \sum_J W_{ij}\, S_j(t) \tag{2}$$

where $V_i(t)$ is the membrane potential of the i-th neuron at t. W_{ij} represent the synaptic weight from the j-th to i-th neuron. $S_j(t)$ represents the spike train of the j-th neuron at time t, indicating whether the neuron spikes (1) or not (0).

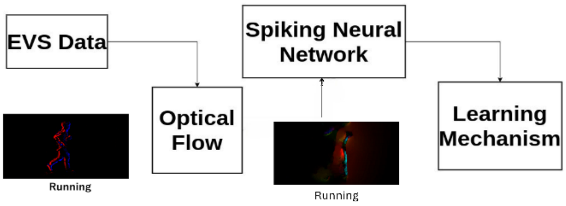


Fig. 2. Proposed system that processes EVS data using the optical flow for feature extraction and performs classification using an SNN.

As shown in Fig. 2, our proposed system processes EVS data by first extracting features through optical flow and then passing the data into the SNN for classification. The architecture is specifically tailored for event-based data. The SNN model processes event-driven data, with key parameters such as the membrane time constant and a spike threshold optimized for performance. The model incorporates decay and reset factors to regulate membrane potential dynamics, ensuring stability and effective signal processing. The learning mechanism involves mapping the preprocessed features to corresponding action labels, with synaptic weight scaling, firing thresholds, and membrane potential decay rates further tuned for optimal results. This integration of optical flow and the SNN ensures robust classification performance for various action recognition tasks. The model architecture consists of an input layer with 1024 neurons, four hidden layers with 512, 256, 128, and 64 neurons, respectively, and an output layer with 4 neurons (one per class).

3. Results and Discussion

We conducted an experiment using a dataset of 600 EVS videos. To ensure robust evaluation, we used 80% of the videos for training and 20% for testing. During training, the network mapped the preprocessed optical flow features to corresponding action labels, with a learning rate of 1×10^{-3} and 10 epochs.

Table 1. Classification table for different type of action.

Action	Precision	Recall	Accuracy
Boxing	1.00	1.00	1.00
Jogging	0.83	0.83	0.83
Running	1.00	0.80	0.90
Walking	0.90	1.00	0.95
Total	0.93	0.90	0.93

The training process, as shown in Fig. 3 demonstrates a steady decline in loss from 1.3 to 0.6 over 10 epochs, indicating that the model is effectively learning and converging towards an optimal solution. The proposed system achieved 93% classification accuracy on the testing set, demonstrating its capability to recognize human action.

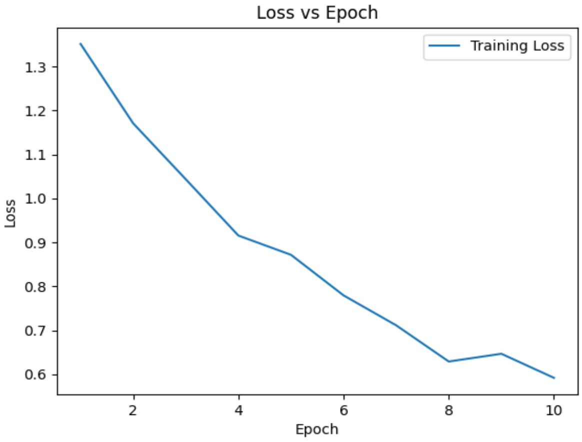


Fig. 3. Training loss over epoch.

As summarized in Table 1, the system has achieved high precision and recall, with boxing and walking performing exceptionally well (1.00 recall for both). While jogging and running exhibit slightly lower recall, their precision remains strong. This combination of efficient training, high accuracy, and strong precision and recall highlights the model's effectiveness for real-time human action recognition.

4. Conclusion

This paper presents an EVS and SNN integrated system for human action classification. By utilizing the event-driven nature of EVS, which only responds to changes in the scene, and the spike-based communication of SNN, we achieved 93% accuracy in classifying human actions. The membrane potential dynamics and spike-based communication between neurons further contributed to the model's effectiveness. For future work, expanding the dataset to include more diverse actions and exploring more advanced SNN architectures, such as hierarchical or recurrent networks, could improve robustness and scalability.

Acknowledgements

This work received support from JSPS KAKENHI Grant Numbers 23H03468, as well as from JST ALCA-Next Grant Number JPMJAN23F3. This work was partially supported by Sony Semiconductor Solutions Corporation.

References

1. Sony Semiconductor Solutions Corporation, “Semiconductor Solutions for Electric Vehicles,” 2024.
2. S. Ghosh-Dastidar and H. Adeli, “Spiking Neural Networks,” Int J Neural Syst, vol. 19, no. 4, pp. 295–308, 2009.
3. Y. Xing, G. Di Caterina, and J. Soraghan, “A New Spiking Convolutional Recurrent Neural Network (SCRNN) with Applications to Event-Based Hand Gesture Recognition,” Front Neurosci, vol. 14, 2020.
4. Q. Liu, D. Xing, H. Tang, D. Ma, and G. Pan, “Event-based Action Recognition Using Motion Information and Spiking Neural Networks,” pp. 1743–1749, 2021.

Authors Introduction

Mr. Aryan Rakheja

He received his bachelor's degree in engineering in 2024 from the Faculty of Computer Science and Engineering, Dronacharya College of Engineering in India. He is currently a master student in Kyushu Institute of Technology, Japan.

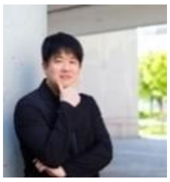
Mr. Rohan Saini

He received his bachelor's degree in engineering in 2024 from the Faculty of Computer Science and Engineering, Dronacharya College of Engineering in India. He is currently a master student in Kyushu Institute of Technology, Japan.

Mr. Ryuta Toyoda

He received the B.Eng. degree from Kyushu Institute of Technology, Japan in 2023. He is a master student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include developing processing circuits for EVS.

Dr. Yuichiro Tanaka



He received the B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology, in 2016, 2018, and 2021, respectively. He was a research fellow with the Japan Society for the Promotion of Science (JSPS), from 2019 to 2021. He was an assistant professor at the Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, from 2021 to 2024. He has been an associate professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology since 2024. His research interests include neural networks, digital hardware implementation, and home service robots. He is a member of IEEE, IEICE, and JNNS.

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.