

Study of a Multi-modal Neurorobotic Prosthetic Arm Control System based on Recurrent Spiking Neural Network

Mark Ikechukwu Ogbodo,^{a)} Khanh N. Dang,^{b)} and Abderazek Ben Abdallah^{c)}

Adaptive Systems Laboratory, Graduate School of Computer Science and Engineering, The University of Aizu, Aizu-Wakamatsu, Fukushima 965-8580, Japan

^{a)}Corresponding author: d8211104@u-aizu.ac.jp

^{b)}Electronic mail: khanh.n.dang@vnu.edu.vn

^{c)}Electronic mail: benab@u-aizu.ac.jp

Abstract.

The use of robotic arms in various fields of human endeavor has increased over the years, and with recent advancements in artificial intelligence enabled by deep learning, they are increasingly being employed in medical applications like assistive robots for paralyzed patients with neurological disorders, welfare robots for the elderly, and prosthesis for amputees. However, robot arms tailored towards such applications are resource-constrained. As a result, deep learning with conventional artificial neural network (ANN) which is often run on GPU with high computational complexity and high power consumption cannot be handled by them. Neuromorphic processors, on the other hand, leverage spiking neural network (SNN) which has been shown to be less computationally complex and consume less power, making them suitable for such applications. Also, most robot arms unlike living agents that combine different sensory data to accurately perform a complex task, use uni-modal data which affects their accuracy. Conversely, multi-modal sensory data has been demonstrated to reach high accuracy and can be employed to achieve high accuracy in such robot arms. This paper presents the study of a multi-modal neurorobotic prosthetic arm control system based on recurrent spiking neural network. The robot arm control system uses multi-modal sensory data from visual (camera) and electromyography sensors, together with spike-based data processing on our previously proposed R-NASH neuromorphic processor to achieve robust accurate control of a robot arm with low power. The evaluation result using both uni-modal and multi-modal input data show that the multi-modal input achieves a more robust performance at 87%, compared to the uni-modal.

Keywords: Multi-modal, Deep Learning, Neurorobotic, Prosthetic Arm, Control System, Electromyography, Spiking Neural Network, Neuromorphic.

INTRODUCTION

The use of robot arms in various areas of human endeavors has increased over the years, with applications ranging from industrial [1, 2], to domestic [3]. In the medical field particularly, robot arms are being employed as assistive robots and prosthetic devices [4, 5] for people with neurological disorders, and amputees. Contemporary robot arms operating in these areas are required to autonomously and adaptively learn new motions and tasks to enable them to operate in a dynamic environment, and in recent years, deep learning has been mostly adopted for this purpose. These requirements, however, bring about new challenges in the area of robot arm control. Deep learning models are based on conventional artificial neural networks (ANN), and although they have shown tremendous performance in cognitive tasks, their tremendous performance requires huge amount of computing power [6] and power consumption [7]; needing much of its complex neural computing strain to be put on graphics processing units (GPUs). This makes them less suitable for the control of resource-constrained autonomous robot arms applied on the edge. As an alternative, neuromorphic systems which have been shown to be less computationally complex than conventional ANN [8] can be employed. Neuromorphic systems implement in hardware, spiking neural network (SNN) which is the third generation of artificial neural networks. Like conventional ANN, SNN is modeled after the structure and computational principles of the brain. However, SNN is more analogous to the brain; communicating via spikes in an event-driven manner. A simple leaky-integrate-and-fire spiking neuron model performs neural computation by first accumulating input spikes at its membrane potential, and when the value of the membrane potential crosses a certain threshold, an output spike is fired. Given a sparse input spike train, the power consumption of SNN is reduced. Neuromorphic systems, therefore, exploit the parallelism and spike sparsity available in SNN to deliver rapid processing with low power [9].

As a prosthetic, a robot arm needs to interact with its surroundings, and to efficiently do this, it needs to go through the steps of sensing, perception, planning, and performing a task. How well a robot arm is able to perform a task in a dynamic space depends on how well it is able to perceive and plan, and this depends on the quality of its sensing. Uni-modal systems have been used in robot control [10]; however, they are sometimes not able to provide sufficient data

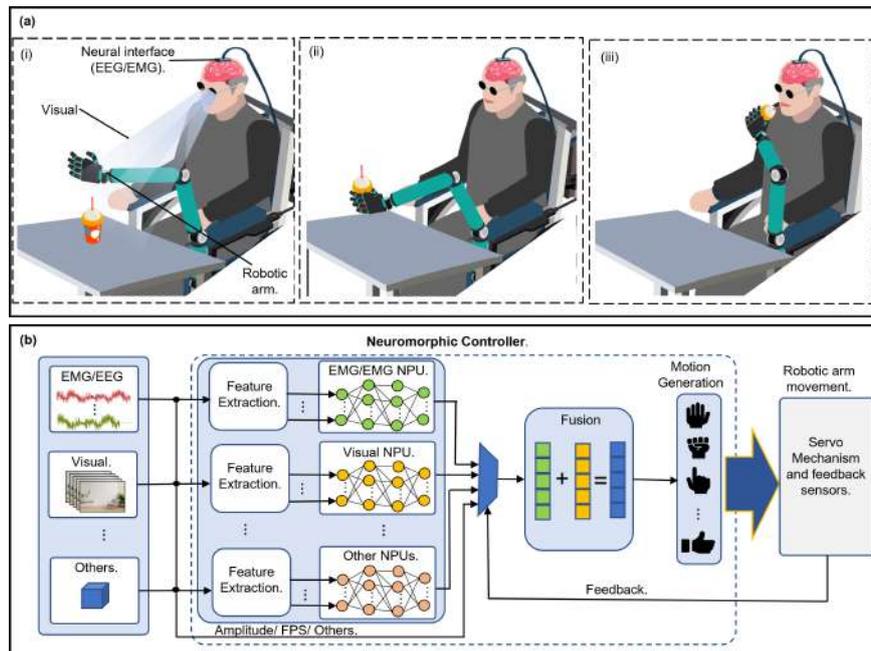


FIGURE 1: High level view of neurorobotic prosthetic arm control system

needed for accurate perception, and this, in turn, prevents the arm from accurately performing a task. The use of multi-modal sensory input, on the other hand, has been shown to achieve more consistent, and accurate perception [11].

While the use of neuromorphic processors in robot control has begun gaining attention, it is important to take into consideration not just how these neuromorphic processors achieve high control accuracy, but also how they deal with faults to ensure reliable and dependable control. The last few decades have seen the extension of robotic applications from the accessible environments of factories and laboratories, to domestic, medical, space, radioactive, and other hazardous environments. The nature of these environments often either makes them least admissible to error due to danger to human lives, or the impracticality of sending humans to make repairs in the event of faults. Faults in robotic control systems have the potential to degrade robot performance or cause fatal accidents, especially in environments like medical, domestic, and space where they are deployed. Therefore, fault-tolerance is crucial in neuromorphic robot control systems to enable them to deal with faults autonomously, and maintain proper operation.

The robot control system is a fundamental part that defines its efficacy, and over the years, several control methods have been employed. Control methods like PID [12, 13] and fuzzy [14, 15] are popularly used; however, the PID controllers are usually not suitable for spatio-temporal systems, and fuzzy being limited to the information encoded within it, cannot learn [16].

Deep learning has shown impressive results as a control method, and advancements in neuromorphic architectures are opening up prospects for addressing some of the technical challenges. In this regard, the application of neuromorphic systems for control in robotics is an active research area and some works have been published. In [17], the authors proposed a spike-based closed-loop control model implemented on two field-programmable gate arrays (FPGA) to control the four joints of a light robot. A neuromorphic robotic control platform was proposed in [18]. The platform was implemented on an unmanned bicycle and adopted a neural state machine that allows the cooperation of hybrid networks to perform specified tasks. In [19], the authors proposed a robot control method based on a mixed-signal analog/digital neuromorphic processor interfaced to a mobile robotic platform that is outfitted with an event-based dynamic vision sensor. A dual-arm aubergine harvesting robot configured to detect and harvest aubergines is proposed in [20]. The authors in [21] presented a human skin-inspired robotic arm ultrasound system for three-dimensional (3D) imaging. To enable efficient and error-free testing of image rectification features on mobile devices, the authors in [22] presented a robotic arm for testing such features.

While these works have explored the use of neuromorphic processors/devices in robot arm control and addressed some underlying issues, there remains a need for multi-modal neuromorphic robot control systems with fault-tolerance for efficient, robust, and reliable robot arm control, especially since none of these works considered fault-tolerance. To

this end, we present in this paper, the study of a multi-modal neurorobotic prosthetic arm control system based on recurrent spiking neural network. The multi-modal neurorobotic arm control system leverages multi-modal sensory data from EMG and visual sensors, with room for more, and a fault-tolerant scalable three-dimensional network on chip (3D-NoC) based digital neuromorphic system with on-chip learning (R-NASH) proposed in our previous work [23, 24] to achieve robust and accurate neurorobotic arm control, in an energy-efficient manner. The contributions of this work are summarised as follows.

- A comprehensive study of a multi-modal neurorobotic prosthetic arm control system based on recurrent spiking neural network.
- A multi-modal fusion algorithm that exploits the complementarity in the modalities of EMG and visual spiking data to achieve precise perception for accurate control.
- Design and evaluation of the proposed multi-modal neurorobotic control system. The performance of the proposed system is validated and used to control a 6 DOF robot arm, while the resource utilization/power analysis is performed on an FPGA prototype of the proposed system based on our previously proposed neuromorphic system R-NASH [23, 24].

SYSTEM ARCHITECTURE AND OVERVIEW

A high-level view of the proposed multi-modal robot arm control system is described in Figure 1. The proposed system takes a modular approach and combines visual and EMG sensory data for precise perception, while enabling scalability for the integration of other modalities of sensory data. Each sensory data is processed by a spike-based neural processing unit of R-NASH (spiking neuron processing core) [24]. The outputs from the neural processing units are either communicated directly to the motion generation unit or fused before sending to the visual generation unit, depending on the control mode. The proposed system implementation is based on our previously proposed robust three-dimensional network on chip (3D-NoC) based neuromorphic system (R-NASH) [23, 24] which provides the energy efficiency and fault-tolerance required for the proposed system in a robust scalable platform. The individual components that make up the proposed system are described below.

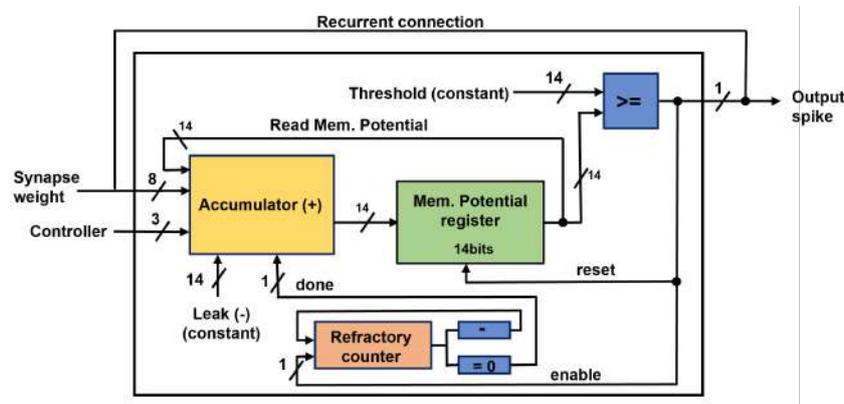


FIGURE 2: Block diagram of a recurrent leaky-integrate-and fire-neuron.

Neural Processing Units

The neural processing units (NPU) are tasked with the initial processing of the data received from various integrated sensors. They are implemented using the leaky-integrate-and-fire (LIF) spiking neuron model for the visual, and a recurrent LIF model shown in Figure 2 for the EMG. The LIF neuron processes spikes by accumulating weighted input spikes in its accumulator and storing the accumulated value as its membrane potential. At the end of each

accumulation time step, a leak operation is performed on the membrane potential and the remainder of its value is compared with a threshold. An output spike is generated if the spike exceeds the threshold, and no spike is generated otherwise. If a spike is generated, the LIF neuron enters a refractory period where it cannot accumulate weighted input spikes until the period is over. The recurrent LIF follows a similar operation, however; unlike the LIF, a recurrent weighted input spike is accumulated at the membrane potential in the event of an output spike.

Visual

The visual NPU implements a feed-forward SNN to classify image frames of hand gestures received from a camera into one of five hand gestures programmed as a command to the robot arm. The captured image frames are transformed to gray-scale, resized, and converted into spikes before being fed to the visual NPU. The recognized gesture is then sent directly to the motion generation unit for visual only control, or the Visual-EMG fusion which fuses its output with that of the EMG NPU before the recognized gesture is communicated to the motion generation unit.

EMG

The EMG NPU implements a recurrent SNN and processes surface electromyography (sEMG) signals from a wearable myoelectric armband that records motor neuron action potentials generated from arm muscle contraction. The sEMG signals are first filtered, then the needed features are extracted, and fed to the EMG NPU to recognize the intended gesture. Similar to the visual NPU, the result of the EMG NPU is also sent to the motion generation unit for sEMG only control or fused with the output from the visual NPU for multi-modal control.

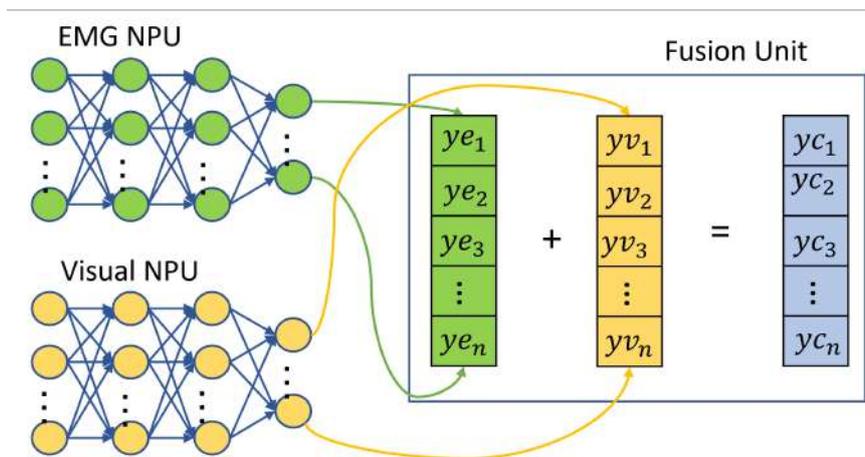


FIGURE 3: Visual-EMG multi-modal fusion

Visual-EMG Multi-modal Fusion

The use of multi-modal data for accurate perception can easily be observed in humans and is employed in control systems to get comprehensive information for control through modalities that describe different aspects of the control environment. As a result, multi-modal data is usually either complementary or supplementary in content, making them more informative than uni-modal data. Early research on the advantages of exploiting multiple data modalities in deep learning can be seen in [25] where the correlation between speech, visual lip motion, and mouth articulations helped improve the accuracy of a speech recognition task. Although humans easily combine different modalities of data from multiple sensory organs to accurately perceive their environment, enabling such ability on a robot control system requires a method of efficiently fusing the data from multiple sensors. Several methods of fusing multi-modal

Algorithm 1: Visual-EMG multi-modal fusion.

```

Input: input visual and EMG spikes (i_vis_spk, i_emg_spk_array)
Output: fused spike (o_fus_spk_array)
1 while true do
    /* valid output spike from visual NPU */
2   if vis_spk_valid == True then
    /* store spikes from visual NPU */
3   vis_spk_reg = i_vis_spk_array;
    /* valid output spike from EMG NPU */
4   if emg_spk_valid == True then
    /* store spikes from EMG NPU */
5   emg_spk_reg = i_emg_spk_array;
6   for i ← 0 to k by 1 do
7   o_fus_spk_array[i] = vis_spk_reg[i] + emg_spk_reg[i];
8   next_state = True;

```

data exist, and they are generally classified either as model-free fusion or model-based fusion. Some examples of the model-free fusion methods include early fusion; which involves concatenating features of the multi-modal data and allowing a classifier to handle the fusion, and late fusion; where uni-modal predictors/classifiers are first used, and multi-modal fusion is performed on their output. The model-based fusion method on the other hand can be achieved using neural networks, graphical models, or kernel methods. In this paper, we employ the model-free late fusion method shown in Figure 3. As described in Algorithm 1, the outputs of the visual and the EMG NPUs are fused for the Visual-EMG fusion unit, where the spikes from each corresponding output neuron of the EMG and visual NPUs; representing the different classes of hand gestures to be recognized, are summed. To determine the correctly recognized gesture, the number of spikes for each summed class is checked and compared, and the class with the highest number of spikes is selected as the recognized gesture. The visual-EMG fusion process can further be described using the following equations.

Fusing the outputs of EMG and Visual NPUs:

$$f(y_{c_i}) = y_{e_i} + y_{v_i} \quad (1)$$

Where y_{e_i} represents the output spikes from each output neuron of the EMG NPU, and y_{v_i} represents the output spikes from each output neuron of the visual NPU.

Determine recognized gesture:

$$\alpha = \max(f(y_{c_1}), f(y_{c_2}), \dots, f(y_{c_n})) \quad (2)$$

Where y_c represents the sum of output spikes from each corresponding EMG and Visual output neurons, and α is the class with the number of spikes.

EVALUATION AND RESULTS

Evaluation Methodology

The proposed system was designed in Verilog-HDL and implemented on the Xilinx Zed board. The synthesis and hardware implementation was performed using Xilinx Vivado, the application that runs on the hardware was designed using the Xilinx SDK, and the bootable files and board configuration were done using Xilinx petalinux. For evaluation, we use a 6 DOF robot arm that consists of 6 bus servos.

As shown in Figure 4 the proposed system was evaluated by classifying sEMG hand gesture dataset from [26] using the EMG NPU, and frame-based hand gesture dataset from a camera using the visual NPU. The output of these NPUs is subsequently fused and also evaluated. A comparison of the performance of the uni-modal and multi-modal control

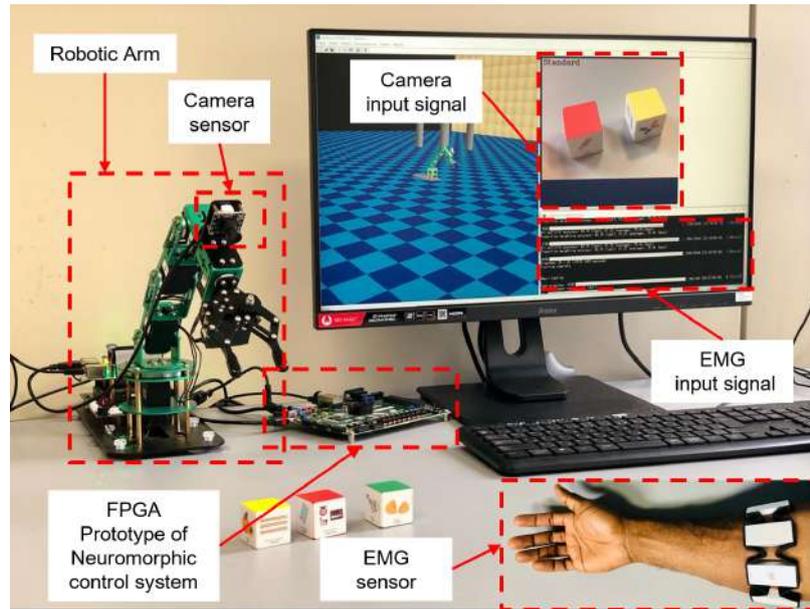


FIGURE 4: Experimental setup for robot arm control.



FIGURE 5: Frame based hand gestures

methods, and also with existing works, are further performed. Each gesture in the datasets is programmed to trigger a set of actions and movements on the robot arm, and depending on the recognized gesture, the robot arm will perform a programmed action.

TABLE I: Network structure and Accuracy of Visual, EMG, and Fusion.

Unit	Type	Configuration	Accuracy
<i>Visual</i>	FC	$256 \times 32 \times 5$	93.1
<i>EMG</i>	SRNN	$16 \times 256 \times 5$	63.98
<i>Fusion</i>	Late	5:5	87.24

Data Acquisition and Preprocessing

As described in Figure 1 the proposed system uses input data from integrated sensors for perception. Currently, two sensors; Visual and sEMG are integrated, which provide hand gestures (pinky, elle, yo, index, and thumb) needed for control in two modalities. The visual data is obtained using a camera that captures RGB image frames of hand gestures. The captured hand gesture image frames are transformed to gray-scale as shown in Figure 5, resized to 16×16 , and converted to spikes using the Poisson encoding scheme [27]. A total of 40,000 (30,000 training, 10,000 testing) was used for the evaluation.

The sEMG follows a similar approach by first capturing raw sEMG signals from an 8 sensor wearable myoelectric armband that records motor neuron action potentials generated from arm muscle contraction. Its dataset consists of 3 sessions of 5 recorded hand gesture signals from 10 subjects, where each gesture recording is about 2 seconds. The recorded sEMG signals sampled at 200Hz are then filtered by first setting the filter cut-off frequencies at 0.5-5Hz. and using a 4th order Butterworth filter with full-wave rectification, and notch (60 Hz). Root mean square (RMS) envelope was performed afterward with a window size of 50ms, and the resulting data was normalized by dividing it with the maximum voluntary contraction (MVC) as shown below.

$$sEMG = \frac{RMS_{envelop}}{MVC} \quad (3)$$

The normalized sEMG values are encoded into spike trains using the Poisson encoding scheme.

Results

Hand Gesture Recognition Performance

To evaluate the classification performance, the Visual, EMG, and Fusion SNN were trained off-chip, and the trained weights were mapped on their respective NPUs for inference. Their accuracy was evaluated individually as shown in Table I. The visual-only uni-modal control achieved an accuracy of 93.3%, the EMG only uni-modal control achieved an accuracy of 63.98%, while the fusion multi-modal control achieved an accuracy of 87%.

TABLE II: Resource utilization of the proposed system.

Resource	Visual	EMG	Fusion	Total
<i>LUT(%)</i>	4	11	1	15
<i>FF(%)</i>	1	2	1	4
<i>BRAM(%)</i>	0	2	0	2
<i>BUFG(%)</i>	1	3	3	7
<i>IO(%)</i>	70	11	3	84
<i>Power(mW)</i>	14	70	1	85

TABLE III: Comparison between the proposed system and existing works.

System	[28]	[29]	[30]	This work
<i>Input signals</i>	EMG	EMG	EMG-DVS	EMG-frame
<i>Gestures</i>	3	3	5	5
<i>Platform</i>	DYNAP	DYNAP	Loihi	FPGA
<i>Accuracy(%)</i>	79.6	83.61	83.4	87.24
<i>Power(mW)</i>	0.05	0.0012	144.5	85

Hardware Utilization

Table II summarizes the hardware utilization of the proposed system on the Xilinx Zedboard. 15% of LUT, 4% of FF, 2% of BRAM, 7% of BUFG, and 84% of IO were utilized. As shown in Figure 6, the visual NPU (neural processing unit) uses 66.09% of these resources. This is because its input size is larger than that of the EMG NPU. The EMG NPU utilizes 26.96% of the total utilized resources following the visual NPU, and finally, the Fusion unit utilizes the least amount of resources at 6.95%.

The power consumption of the proposed system is also presented in Table II. At 100MHz, the EMG NPU consumes the most power at 70mW, followed by the visual NPU, at 14mW. This disparity in power consumption between the

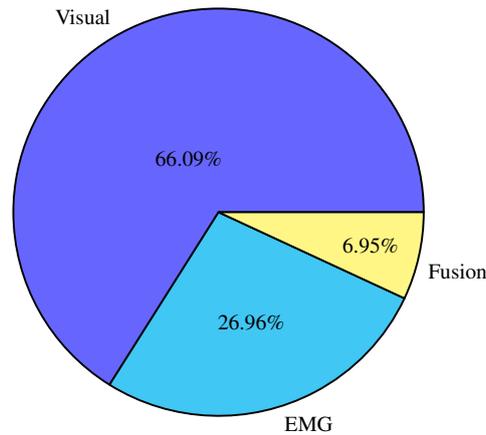


FIGURE 6: Summary of resource utilization.

visual NPU and the EMG NPU is due to the larger number of hidden layer neurons required to process the sEMG signals, which takes more processing operations compared to that of the visual. The fusion unit consumes the least power at 1mW because the fusion process does not require as many inputs and processing as the other units.

Result Comparison with Existing Works

A comparison of the proposed system with existing works is presented in Table III, and as shown, the works in [29] and [28] achieved higher accuracy, and lower power consumption. However, they performed uni-modal recognition of only 3 sEMG hand gestures. The work in [30] performed multi-modal recognition of hand gestures similar to the proposed system, with the difference being in the use of dynamic vision sensor (DVS) data in [30], and frame-based data in the proposed work. Nevertheless, the proposed work achieves higher accuracy, and lower power consumption.

DISCUSSION

In previous sections, we presented the proposed multi-modal neurobotic prosthetic arm control system based on recurrent spiking neural networks and described its architecture and design. We also evaluated its performance and presented a comparison of uni-modal and multi-modal control results, and that of existing works. In light of this, we discuss in this section, issues related to the evaluation and results.

An issue we observed is that a lengthier sEMG signal period with time steps of around 150 is needed to achieve considerable accuracy for the EMG control, while the visual control requires about 40 Poisson-generated time steps. When using the uni-modal control method this does not affect their original performance, however; when fusing, an equal number of time steps are required to ensure there is no bias; where a wrong gesture is recognized because EMG NPU generated more spikes for that class due to its increased time steps, to overwhelm the impact of the visual NPU spikes. With an equal number of time steps, the early classification feature of SNN could easily be achieved, where the fused spikes are transmitted each time step, with no need to wait for all time steps to get a sense of the probable recognized gesture. Nevertheless, increasing the number of time steps required for the visual NPU to match that of the EMG leads to unnecessary operations which will not impact its performance, and this leads to unnecessary energy consumption. Reducing the number of time steps of the sEMG signal, on the other hand, reduces the performance of the EMG NPU. A suitable ratio-based approach applied to both data could mitigate the impact of the difference in time steps.

CONCLUSION AND FUTURE WORKS

In this paper, we presented a multi-modal neurobotic arm control system based on recurrent spiking neural network prototyped on a Zedboard FPGA based on our previously proposed neuromorphic system R-NASH [23, 24]. The proposed system enables the control of neurobotic arm using multi-modal sensory data from both visual and sEMG sensors while enabling a scalable integration of other sensory data modalities. The evaluation result shows that uni-modal control with frame-based visual data, and sEMG data, achieves accuracies of 93.1% and 63.96% respectively, while multi-modal control combines both EMG and visual control to achieve an accuracy of 87%. Subsequent work will focus on improving the EMG NPU performance, recognition of more gestures, and integration of more sensors for better perception.

REFERENCES

1. S. Grahn, B. Langbeck, K. Johansen, and B. Backman, "Potential advantages using large anthropomorphic robots in human-robot collaborative, hand guided assembly," *Procedia CIRP* **44**, 281–286 (2016).
2. A. Robotics, "Collaborative robots," <https://new.abb.com/products/robotics/collaborative-robots> (2010), accessed: 2021-10-08.
3. M. Ababneh, H. Sha'ban, D. AlShalabe, D. Khader, H. Mahameed, and M. AlQudimat, "Gesture controlled mobile robotic arm for elderly and wheelchair people assistance using kinect sensor," in *2018 15th International Multi-Conference on Systems, Signals Devices (SSD)* (2018) pp. 636–641.
4. K. Li, Y. Fang, Y. Zhou, and H. Liu, "Non-invasive stimulation-based tactile sensation for upper-extremity prosthesis: A review," *IEEE Sensors Journal* **17**, 2625–2635 (2017).
5. O. Bionics, "Turning disabilities into superpowers," <https://openbionics.com/> (2021), accessed: 2021-10-08.
6. R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Communications of the ACM* **63**, 54–63 (2020).
7. E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," *Computer Research Repository abs/1906.02243* (2019), 1906.02243.
8. C. Lee, S. Sarwar, P. Panda, G. Srinivasan, and K. Roy, "Enabling spike-based backpropagation for training deep neural network architectures," *Frontiers in Neuroscience* **14**, 119 (2020).
9. M. Ogbodo, T. Vu, K. Dang, and A. Abdallah, "Light-weight spiking neuron processing core for large-scale 3D-NoC based spiking neural network processing systems," in *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)* (Institute of Electrical and Electronics Engineers (IEEE), 2020) pp. 133–139.
10. R. A. de Lemos, L. A. C. de O. Nogueira, A. M. Ribeiro, L. G. B. Mirisola, M. F. Koyama, E. C. de Paiva, and S. S. Bueno, "Unisensory intra-row navigation strategy for orchards environment based on sensor laser," (SBA Sociedade Brasileira de Automática, 2018).
11. W. Guo, J. Wang, and S. Wang, "Deep multimodal representation learning: A survey," *IEEE Access* **7**, 63373–63394 (2019).
12. J. Lee, P. H. Chang, B. Yu, and M. Jin, "An adaptive PID control for robot manipulators under substantial payload variations," *IEEE Access* **8**, 162261–162270 (2020).
13. S. Zhen, Z. Zhao, X. Liu, F. Chen, H. Zhao, and Y.-H. Chen, "A novel practical robust control inheriting PID for scara robot," *IEEE Access* **8**, 227409–227419 (2020).
14. C. Sun, C. Liu, X. Feng, and X. Jiao, "Visual servoing of flying robot based on fuzzy adaptive linear active disturbance rejection control," *IEEE Transactions on Circuits and Systems II: Express Briefs* **68**, 2558–2562 (2021).
15. R. H. Abiyev, N. Akkaya, and I. Günsel, "Control of omnidirectional robot using z-number-based fuzzy system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **49**, 238–252 (2019).
16. P. B. and, "Robotics controller: A literature survey," *International Journal of Engineering Research and Technology* **4**, 1–7 (2015).
17. A. Linares-Barranco, F. Perez-Peña, A. Jimenez-Fernandez, and E. Chicca, "ED-BioRob: A neuromorphic robotic arm with fpga-based infrastructure for bio-inspired spiking motor controllers," *Frontiers in Neurobotics* **14**, 96 (2020).
18. Z. Zou, R. Zhao, Y. Wu, Z. Yang, L. Tian, S. Wu, G. Wang, Y. Yu, Q. Zhao, M. Chen, J. Pei, F. Chen, Y. Zhang, S. Song, M. Zhao, and L. Shi, "A hybrid and scalable brain-inspired robotic platform," (Springer Science and Business Media LLC, 2020) p. 18160.
19. H. Blum, A. Dietmüller, M. Milde, J. Conradt, G. Indiveri, and Y. Sandamirskaya, "A neuromorphic controller for a robotic vehicle equipped with a dynamic vision sensor," (2017) pp. 1–9.
20. D. Sepúlveda, R. Fernández, E. Navas, M. Armada, and P. González-De-Santos, "Robotic aubergine harvesting using dual-arm manipulation," *IEEE Access* **8**, 121889–121904 (2020).
21. Q. Huang, J. Lan, and X. Li, "Robotic arm based automatic ultrasound scanning for three-dimensional imaging," *IEEE Transactions on Industrial Informatics* **15**, 1173–1182 (2019).
22. D. Banerjee, K. Yu, and G. Aggarwal, "Image rectification software test automation using a robotic arm," *IEEE Access* **6**, 34075–34085 (2018).
23. A. Ben Abdallah and K. N. Dang, "Toward robust cognitive 3D brain-inspired cross-paradigm system," *Frontiers in Neuroscience* **15**, 795 (2021).
24. O. M. Ikechukwu, K. N. Dang, and A. B. Abdallah, "On the design of a fault-tolerant scalable three dimensional NoC-based digital neuro-morphic system with on-chip learning," *IEEE Access* **9**, 64331–64345 (2021).
25. M. HARRY and M. JOHN, "Hearing lips and seeing voices," *Nature* **264**, 746–748 (1976).
26. E. Ceolini, G. Taverni, M. Payvand, and E. Donati, "Emg and video dataset for sensor fusion based hand gestures recognition | zenodo," <https://zenodo.org/record/3228846#.Yip1ZnpBxhF> (2019), (Accessed on 03/11/2022).

27. M. A. Triplett and G. J. Goodhill, "Probabilistic encoding models for multivariate neural data," *Frontiers in Neural Circuits* **13** (2019), 10.3389/fncir.2019.00001.
28. E. Donati, M. Payvand, N. Risi, R. Krause, and G. Indiveri, "Discrimination of EMG signals using a neuromorphic implementation of a spiking neural network," *IEEE Transactions on Biomedical Circuits and Systems* **13**, 795–803 (2019).
29. Y. Ma, B. Chen, P. Ren, N. Zheng, G. Indiveri, and E. Donati, "EMG-based gestures classification using a mixed-signal neuromorphic processing system," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **10**, 578–587 (2020).
30. E. Ceolini, C. Frenkel, S. B. Shrestha, G. Taverni, L. Khacef, M. Payvand, and E. Donati, "Hand-gesture recognition based on EMG and event-based camera sensor fusion: A benchmark in neuromorphic computing," *Frontiers in Neuroscience* **14** (2020), 10.3389/fnins.2020.00637.