



## OPEN ACCESS

## EDITED BY

Mariel Alfaro-Ponce,  
Monterrey Institute of Technology and  
Higher Education (ITESM), Mexico

## REVIEWED BY

Raymundo Cassani,  
McGill University Health Centre, Canada  
Daniela Diaz Alonso,  
Center for Engineering and  
Development CIDESI, Mexico

## \*CORRESPONDENCE

Giovanny Sanchez,  
gsanchezriv@ipn.mx  
Juan-Gerardo Avalos,  
javaloso@ipn.mx

## SPECIALTY SECTION

This article was submitted to Biomedical  
Robotics,  
a section of the journal  
Frontiers in Robotics and AI

RECEIVED 25 August 2022

ACCEPTED 08 September 2022

PUBLISHED 23 September 2022

## CITATION

Anides E, Garcia L, Sanchez G,  
Avalos J-G, Abarca M, Frias T,  
Vazquez E, Juarez E, Trejo C and  
Hernandez D (2022), A biologically  
inspired spiking neural P system in  
selective visual attention for efficient  
feature extraction from human motion.  
*Front. Robot. AI* 9:1028271.  
doi: 10.3389/frobt.2022.1028271

## COPYRIGHT

© 2022 Anides, Garcia, Sanchez, Avalos,  
Abarca, Frias, Vazquez, Juarez, Trejo  
and Hernandez. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# A biologically inspired spiking neural P system in selective visual attention for efficient feature extraction from human motion

Esteban Anides<sup>1</sup>, Luis Garcia<sup>1</sup>, Giovanny Sanchez<sup>1\*</sup>,  
Juan-Gerardo Avalos<sup>1\*</sup>, Marco Abarca<sup>1</sup>, Thania Frias<sup>1</sup>,  
Eduardo Vazquez<sup>1</sup>, Emmanuel Juarez<sup>2</sup>, Carlos Trejo<sup>2</sup> and  
Derlis Hernandez<sup>2</sup>

<sup>1</sup>Instituto Politecnico Nacional ESIME Culhuacan, Mexico City, Mexico, <sup>2</sup>Tecnologico Nacional de Mexico, Tecnologico de Estudios Superiores de Ecatepec, Estado de Mexico, Mexico

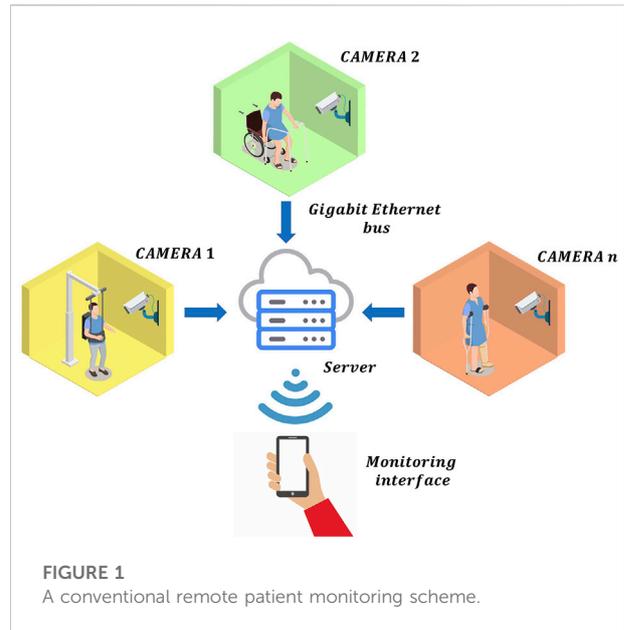
Nowadays, human action recognition has become an essential task in health care and other fields. During the last decade, several authors have developed algorithms for human activity detection and recognition by exploiting at the maximum the high-performance computing devices to improve the quality and efficiency of their results. However, in real-time and practical human action recognition applications, the simulation of these algorithms exceed the capacity of current computer systems by considering several factors, such as camera movement, complex scene and occlusion. One potential solution to decrease the computational complexity in the human action detection and recognition can be found in the nature of the human visual perception. Specifically, this process is called selective visual attention. Inspired by this neural phenomena, we propose for the first time a spiking neural P system for efficient feature extraction from human motion. Specifically, we propose this neural structure to carry out a pre-processing stage since many studies have revealed that an analysis of visual information of the human brain proceeds in a sequence of operations, in which each one is applied to a specific location or locations. In this way, this specialized processing have allowed to focus the recognition of the objects in a simpler manner. To create a compact and high speed spiking neural P system, we use their cutting-edge variants, such as rules on the synapses, communication on request and astrocyte-like control. Our results have demonstrated that the use of the proposed neural P system increases significantly the performance of low-computational complexity neural classifiers up to more 97% in the human action recognition.

## KEYWORDS

spiking neural P systems, selective visual attention, human action detection, human visual perception, communication on request, astrocyte-like control, rules on the synapses

## Introduction

Nowadays, human action recognition (HAR) has become a vital task in many applications, such as intelligent human–machine interfaces, intelligent video surveillance, video storage and retrieval and home monitoring (Liu J. et al., 2019). Specifically, remote patient monitoring (RPM) has emerged as a potential solution for rehabilitation purposes since the number of patients with stroke or other disability has increased significantly. Therefore, human action recognition plays an important role in this application since the movement of patients needs to be continuously monitored and rectified. In this way, their motion patterns can be corrected. Nowadays, current IoT technologies, such as RPM, glucose monitoring, heart-rate monitoring, depression and mood monitoring, among others, have allowed to the doctors take care of the health of their patients from home (Liu Z. et al., 2019; Bouchabou et al., 2021). In particular, RPM technology have allowed to therapists monitor the patients, who are in a home-based rehabilitation scheme, remotely by using broadband internet (Hasan et al., 2019; Farias et al., 2020). Until date, many algorithms have developed to detect and recognize the human motion under controlled environments to achieve high accuracy. Despite this, the simulation of these advanced algorithms overcomes the computational capacity of the current computing systems. In addition, bandwidth congestion may occur when trying to send multiple videos to the doctor. As a consequence, the detection and recognize of the human motion in real-time becomes infeasible since the latency of the RPM system is increased significantly. In the past decades, several approaches have been developed to efficiently perform human action recognition (Baccouche et al., 2011; Ji et al., 2012; Grushin et al., 2013; Ali and Wang, 2014; Liu et al., 2014; Shu et al., 2014; Shi et al., 2015; Veeriah et al., 2015; Dash et al., 2021). In general terms, these approaches intend to determine when the action is occurring and what is this action by considering the starting and ending times of all action occurrences from the video. Despite achieving these important developments, there are still several challenges to overcome. In particular, most of these approaches show high performance in terms of recognition accuracy at the cost of exhibiting high computational complexity. As a consequence, the use of these approaches in practical real-time RPM application can be limited. Therefore, there is still the need to develop low-computational complexity algorithms with high recognition accuracy. One potential solution can be found in the nature. In particular, in the human vision in which a compact representation of the visual content in the form of salient objects/events can be obtained. Recently, advanced approaches have been developed based on the visual attention. For example, (Sharma et al., 2015) presented a



visual attention algorithm based on multi-layered recurrent neural networks (RNNs) to efficiently perform action recognition in videos. However, the model exhibits high computational complexity by performing a large number of convolutional operations. Other example can be found in Guo et al. (2022). The authors developed a linear attention method named large kernel attention (LKA) for visual tasks. In general terms, these works intend to improve the recognition accuracy by increasing their computational complexity. From the engineering perspective, this factor is vital since most of the current processing devices are portable. Therefore, the simulation of these approaches exceeds the computational capabilities of these devices. Here, we propose for the first time a method to create an efficient pre-processing step inspired by the visual attention mechanism. In this way, any classification scheme can be used to improve their performance significantly. In this way, the human action recognition can be performed efficiently by ensuring a low computational complexity.

## Materials and methods

### The proposed pre-processing visual system based on spiking neural P systems

Nowadays, several patients can be monitored remotely by using the current IoT technologies, as shown in Figure 1. Under this RPM scheme, the information of each patient is sent to the doctor via Gigabit Ethernet. In general, this information is composed of frames of the whole scene with ultra-high

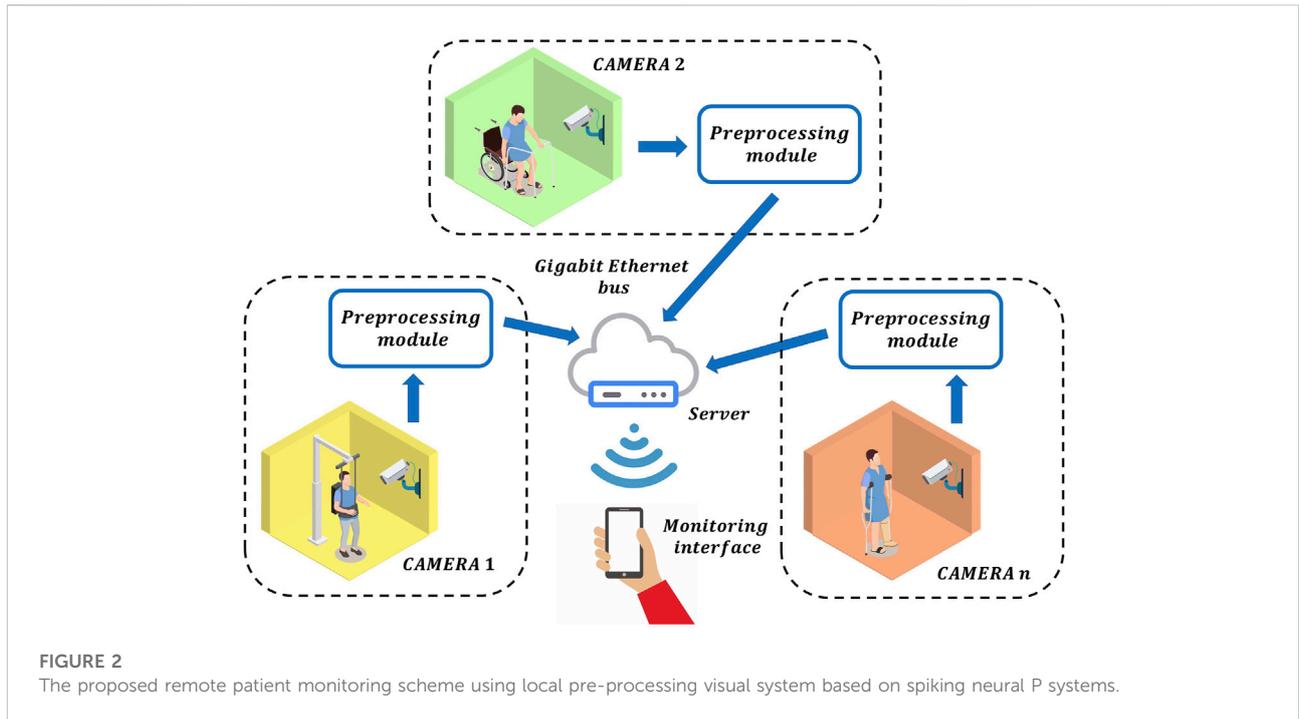


FIGURE 2 The proposed remote patient monitoring scheme using local pre-processing visual system based on spiking neural P systems.

definition, which implies to send billions of bits at each second per each patient to the server. Therefore, a bandwidth congestion can occur by monitoring multiple patients. Obviously, this information can be compressed by using different video data compression methods. However, the latency of the RPM scheme is increased significantly by performing this additional process. Therefore, the monitoring of the patients in real-time is unfeasible. To avoid this, one potential solution can be found if each local monitoring system only sends the relevant information to detect the movement of the patients instead of sending the whole scene. Inspired by the selective attention, we propose a model of pre-processing before data analysis and learning. This neural behaviour can be found in the human visual system since this obtains a compact representation of the visual content in the form of salient events/objects. In general, when the visual system detects a change in position or pose in an observed object, it produces a different pattern of activity from a population of neurons, corresponding to a different visual association response (DiCarlo et al. (2012)). By exploiting these advantages to the maximum, we propose a RPM scheme using local pre-processing visual system based on spiking neural P (SN P) systems, as shown in Figure 2. In this system, the pre-processing stage can be performed by an embedded system, which only sends the relevant information to the doctor in real-time. The SN P systems were proposed as a new class of distributed and parallel systems (Ionescu et al. (2006)). Nowadays, some authors have proposed novel variants of SN P systems such as weighted synapses (Pan et al. (2012)), with communication on request (Pan et al. (2017)) and rules on

synapses (Song et al. (2014)). These systems have had a great impact, since they have proven to be useful in engineering applications.

Before presenting the proposed remote patient monitoring scheme, we provide the formal definition of SN P systems ( $\Pi$ ) of degree  $m \geq 1$ , as follows:

$$\Pi = \{O, \sigma_1, \sigma_2, \dots, \sigma_m, \text{syn}, \text{in}, \text{out}\}$$

Where

- 1)  $O = \{a\}$  is the singleton alphabet, where  $a$  is called *spike*;
- 2)  $\sigma_1, \sigma_2, \dots, \sigma_m$  are neurons of the form  $\sigma_i = (n_i, R_i)$

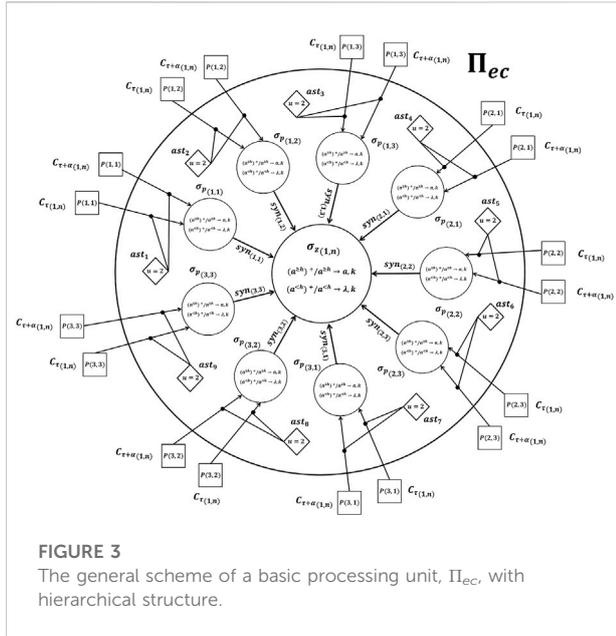
where

- a)  $n_i \geq 0$  is the initial number of spikes contained in neuron  $\sigma_i$ ;
- b)  $R_i$  is the finite set of rules of the following two forms:
  - (1)  $E/a^c \rightarrow a$ , where  $E$  is a regular expression over alphabet  $O$ ,  $c \geq 1$ ;
  - (2)  $a^s \rightarrow \lambda$ , for  $s \geq 1$  and  $a^s \notin L(E)$ , where  $\lambda$  represents the empty string;
- 3)  $\text{syn}$  is a set of synapses, where each element is given by:

$$(i, j) \subseteq \{1, \dots, m\} \times \{1, \dots, m\}, \text{Where}$$

- a)  $m$  is the total number of neurons ( $\sigma$ );
- b)  $i, j$  denote pre-synaptic and post-synaptic neurons, respectively ( $1 \leq i, j \leq m$ ).

Here, synapses are defined as follows:



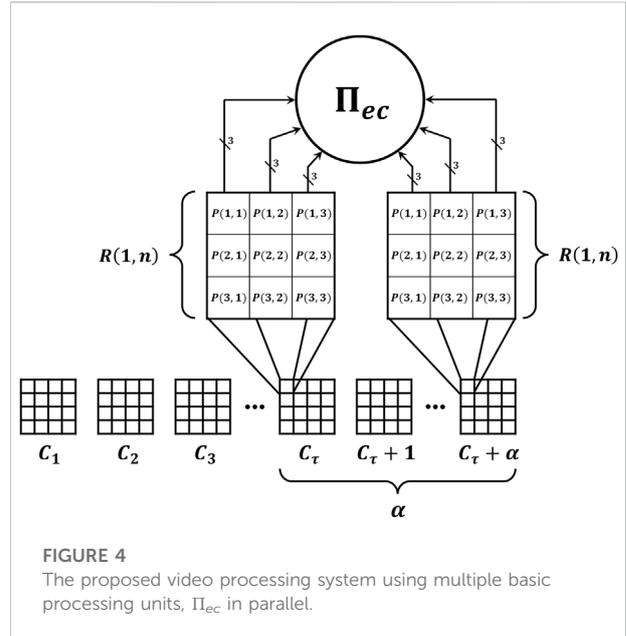
**FIGURE 3**  
The general scheme of a basic processing unit,  $\Pi_{ec}$ , with hierarchical structure.

$$syn = (\sigma_i, \sigma_j, \Delta t_d, R)$$

- a)  $\Delta t_d$  defines the maximum number of steps required for a spike to reach neuron  $j$  via neuron  $i$  (Garcia et al. (2021)).
- b)  $R_{(i,j)}$  is a finite set of rules of the following two forms:
  - $E/a^c \rightarrow a^p; d$  (firing rule), where  $E$  is a regular expression over  $O$ ,  $c \geq p \geq 1$  and  $d \geq 0$ .
  - $a^s \rightarrow \lambda$  (forgetting rule), for some  $s \geq 1$ , with the restriction that  $a^s \notin L(E)$  for any rule  $E/a^c \rightarrow a^p; d$  from any  $R_{(i,j)}$ .
- 4)  $in, out \in \{1, \dots, m\}$  indicates the input and the output neurons, respectively.

Once the formal definition was provided, we present the structure of the spiking neural P system, as shown in Figure 3. Here, we take inspiration of the neural activity at all levels of the visual system from the retina to regions of parietal and frontal cortex. Specifically, neurons in early visual areas have small spatial receptive fields (RFs), and neurons in later areas have large RFs and code abstract features such as behavioral relevance. In this way, the information is organized to obtain local and global features of a image. Therefore, selective attention coordinates the activity of neurons to compact representations of an object. Inspired by this neural behaviour, we build a SN P system with two layers. The first layer, which is composed of neurons,  $\sigma_p$ , intends to mimic the behaviour of the small spatial receptive fields, and the second layer is composed of a set of neurons,  $\sigma_{z(m,m)}$ .

Based on the hierarchical structure of the proposed processing unit,  $\Pi_{ec}$ , the definition can be expressed as follows:

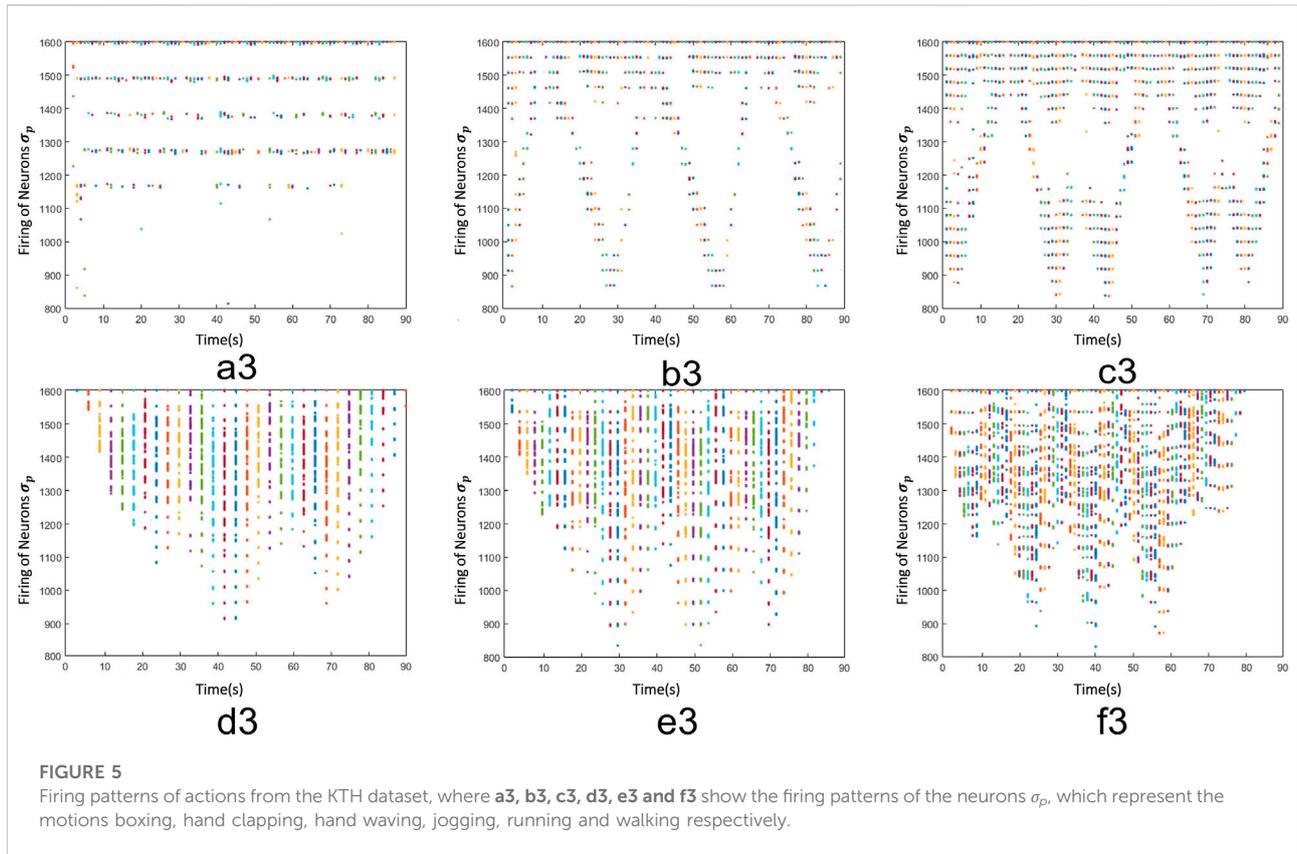


**FIGURE 4**  
The proposed video processing system using multiple basic processing units,  $\Pi_{ec}$  in parallel.

$$\begin{aligned} \Pi_{ec} &= \{0, \sigma_z, \sigma_{p(1,1)}, \dots, \sigma_{p(3,3)}, ast_1, \dots, ast_9, in, out\} \\ 0 &= \{a\} \\ \sigma_{p_{i,j}} | (i, j) \in \{(1, 1), (1, 2), \dots, (n, m)\} &= (0, R_{p_1}, R_{p_2}) \\ \sigma_z &= (0, R_{z_1}, R_{z_2}) \\ R_{p_1} &= \{(a^{zh})^+ / a^{zh} \rightarrow a, k\}, \\ R_{p_2} &= \{(a^{ch})^+ / a^{ch} \rightarrow \lambda, k\}, \\ R_{z_1} &= \{(a^{zh'})^+ / a^{zh'} \rightarrow a, k\}, \\ R_{z_2} &= \{(a^{ch'})^+ / a^{ch'} \rightarrow \lambda, k\} \\ in &= \{P_{(1,1)}, P_{(1,2)}, \dots, P_{(n,m)}\} \\ out &= \{\sigma_z\} \\ ast_i | i \in \{0, 1, \dots, 9\} &= \{u = 2, syn_{i_1}, syn_{i_2}\} \end{aligned}$$

The proposed processing unit,  $\Pi_{ec}$ , with hierarchical structure works as follows:

Initially, the user must define the distance between several frames by means of the factor,  $\alpha$ , and the number of regions,  $R$ , per each frame,  $C$ , as shown in Figure 4. Specifically, the distance,  $\alpha$ , between frames,  $C$ , defines the speed at which the video is analyzed. On the other hand, the regions define the area of analysis within each frame,  $C$ . Therefore, each region of a frame,  $C$ , represents a part of the human body. Once the parameters,  $\alpha$ , and  $R$  were defined, the value of each pixel is converted into its equivalent number of spikes. For example, if the pixel value is equal to 5, this value is represented by means of a train of five spikes. Once the pixels are converted into spikes, these are sent to the first layer of neurons,  $\sigma_p$ . As can be seen in Figure 3, neurons,  $\sigma_p$ , have astrocytes,  $ast$ , to regulate the input spikes. Here, these astrocytes performs an intrinsic subtraction by means of their inhibition rule. Otherwise, only one spike is sent to its respective neuron,  $\sigma_p$ . The use of this mechanism has allowed us to perform the comparison between two frames,  $C$ . In this way, the movement of the human can be detected, i.e., if there is a difference between two consecutive frames, there is significant



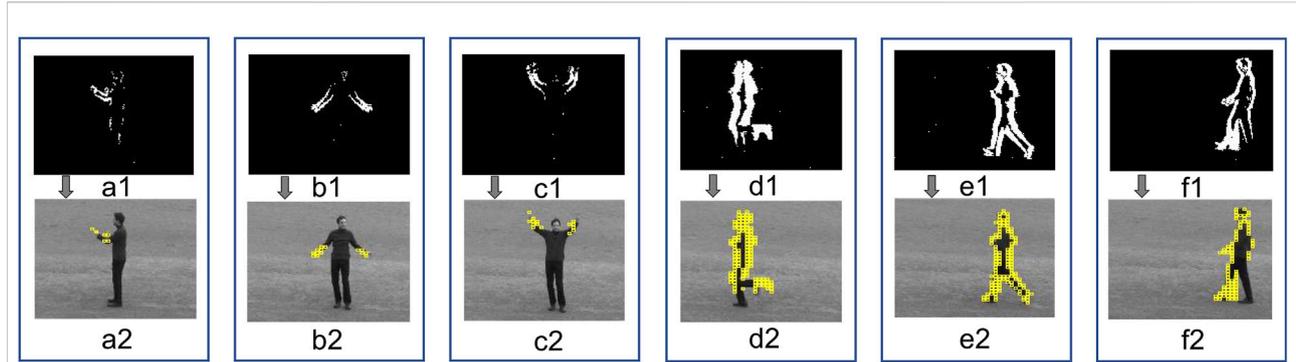
movement of the human. Therefore, only significant information can be processed by neurons,  $\sigma_p$  and the remaining information is discarded. Once the input pulses are delivered to neurons,  $\sigma_{p(i,j)}$ , these neurons fire the accumulated spikes to neuron,  $\sigma_z$ . In general, neurons,  $\sigma_z$  were proposed to indicate the motion of a particular region,  $R$ , by applying their firing rules,  $((a^{\geq h})^+ / a^{\geq h} \rightarrow a, k)$ , where  $k$  is in function of the bit-resolution of the image, and  $h$  is a threshold, which is defined in terms of the number of pixels per each region,  $R$ . Therefore, this threshold is used as descriptor to indicate the information of the spatial location of regions,  $R$ , with significant movement.

## Results

To demonstrate the performance of the proposed processing units,  $\Pi_{ec}$ , we use the KTH dataset (Schuldt et al., 2004). In this way, we make a coherent comparison between our approach and existing works (Baccouche et al., 2011; Ji et al., 2012; Grushin et al., 2013; Ali and Wang, 2014; Liu et al., 2014; Shu et al., 2014; Shi et al., 2015; Veeriah et al., 2015; Dash et al., 2021). This data set is composed of 6 videos, which involves human actions, such as boxing(a), hand clapping(b), hand waving(c), jogging(e), running (d) and walking(f). These activities were carried out by 25 people under four different conditions, such as outdoors,

outdoors with different clothes, outdoors with scale variation and indoors. In addition, these videos have a resolution of  $120 \times 160$  pixels. To perform our experiments, each frame,  $C$ , is divided into 1,600 regions, in which each region is composed of  $3 \times 4$  pixels. In this way, the padding is avoided. Furthermore, we select the thresholds  $h = 15$ ,  $h' = 6$  and  $\alpha = 6$  by analyzing the speed of different human motions, as shown in Figure 5. Specifically, the proposed processing units,  $\Pi_{ec}$ , computes 6 frames at each 240 ms since the video is composed of 25 frames per second. At this rate, processing units,  $\Pi_{ec}$ , generate diverse firing patterns, which contain spatial and temporal information of each action. Here, we used 216 videos, which correspond to 9 people carrying out the 6 different activities in 4 different environments, 50% was used for training and the other 50% for testing and validation. Obviously, regions with significant motion (d, e and f) generate firing patterns with high number of spikes, otherwise human motions (a, b and c) produce a fewer number of spikes. Here, these patterns have helped us to create characteristic vectors.

Our preliminary results show that the processing units,  $\Pi_{ec}$ , are able to identify the human motions, as shown in Figure 6. As can be observed, images (a1, b1, c1, d1, e1 and f1) show the regions,  $R$ , in which the human motion is detected by the proposed processing units,  $\Pi_{ec}$  and images (a2, b2, c2, d2, e2 and f2) show which part of the human body motion moves.



**FIGURE 6**  
 Detection of the human motion using the proposed processing units,  $\Pi_{ec}$ , where **a1, b1, c1, d1, e1, f1** show the regions  $R$  of the motions boxing, hand clapping, hand waving, jogging, running and walking respectively, additionally **a2 b2, c2, d2, e2, f2** show which part of the human body represents the motions boxing, hand clapping, hand waving, jogging, running and walking respectively.

**TABLE 1** Performance comparison between different classifiers of the classification learner tool.

Classifier	Average Performance (%) without using $\Pi_{ec}$	Average Performance (%) using $\Pi_{ec}$
SVM	83%	97.2%
WKNN	78%	97.2%
ESD	85%	97.7%
EBT	80%	98.1%

**TABLE 2** Performance comparison between the proposed human action recognition system and existing algorithms.

Author	Algorithm	Dataset	Real time	Performance (%)
This work	SNP and SVM	KTH (Schuldt et al., 2004)	Yes	97.2%
Dash et al., (2021)	SIFT and CNN	KTH(Schuldt et al., 2004)	No	90.0%
Veeriah et al., (2015)	Differential RNN	KTH(Schuldt et al., 2004)	No	93.9%
Shi et al., (2015)	DTT, DNN	KTH (Schuldt et al., 2004)	No	95.6%
Liu et al., (2014)	MTSL	KTH (Schuldt et al., 2004)	No	96.7%
Ali and Wang, (2014)	DBN and SVM	KTH (Schuldt et al., 2004)	No	94.3%
Shu et al., (2014)	3DGab and SNN	KTH (Schuldt et al., 2004)	No	92.3%
Grushin et al., (2013)	LSTM	KTH (Schuldt et al., 2004)	No	90.7%
Ji et al., (2012)	3DCNN	KTH (Schuldt et al., 2004)	No	90.02%
Baccouche et al., (2011)	CNN and RNN	KTH (Schuldt et al., 2004)	No	94.3%

Once the characteristic vectors were defined, these are used to classify different human actions. To classify these actions, we use the tool called “classification learner” of Matlab, which contains several classifiers, such as Support Vector Machine (SVM), Weighted k nearest neighbors (WKNN), Ensemble Subspace Discriminant (ESD), and Ensemble Bagged Trees (EBT). It should be noted that this tool contains low complexity

classifiers. As a consequence, their performance in terms of the recognition accuracy is limited, as shown in Table 1. As can be observed, the use of the proposed processing units,  $\Pi_{ec}$ , significantly increases the performance of these low-computational complexity classifiers. Specifically, the hierarchical organization of the proposed SNP system has shown that selective attention coordinates the neural activity

to detect objects in a compact visual representation. In addition, we compare the performance of our approach and existing works, as shown in Table 2. As can be observed, the proposed human action recognition achieves the best performance reported to date. It should be noted that the existing classifiers intend to achieve high performance at the cost of increasing their computational complexity (Ghamisi et al., 2017). In our approach, we use one of the most simplest algorithm along with the proposed processing units,  $\Pi_{ec}$  to significantly improve the performance of the human action recognition system.

## Conclusion

In this work, we present for the first time a pre-processing method based on the spiking neural systems for efficient feature extraction in video sequences for motion detection of human actions. Specifically, we propose a basic processing unit,  $\Pi_{ec}$ , with hierarchical structure inspired by the selective attention mechanism. Since the SN P systems exhibits an intrinsic parallel processing, we propose a parallel video processing system by using multiple processing units,  $\Pi_{ec}$ , where each unit is implemented by mainly using comparators. As a consequence, the proposed units exhibit low computational complexity. From the engineering point of view, this potentially allows their implementation in embedded devices, where the area is limited. Part of future work is to perform more experiments by building a remote patient monitoring.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## References

- Ali, K. H., and Wang, T. (2014). "Learning features for action recognition and identity with deep belief networks," in 2014 International Conference on Audio, Language and Image Processing, New Delhi, India, October 5–8, 2009. (IEEE), 129.
- Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., and Baskurt, A. (2011). "Sequential deep learning for human action recognition," in *International workshop on human behavior understanding* (Springer), 29.
- Bouchabou, D., Nguyen, S. M., Lohr, C., LeDuc, B., and Kanellos, I. (2021). A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. *Sensors* 21, 6037. doi:10.3390/s21186037
- Dash, S. C. B., Mishra, S. R., Srujan Raju, K., and Narasimha Prasad, L. (2021). Human action recognition using a hybrid deep learning heuristic. *Soft Comput.* 25, 13079–13092. doi:10.1007/s00500-021-06149-7
- DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi:10.1016/j.neuron.2012.01.010
- Farias, F. A. C. d., Dagostini, C. M., Bicca, Y. d. A., Falavigna, V. F., and Falavigna, A. (2020). Remote patient monitoring: A systematic review. *Telemedicine e-Health* 26, 576–583. doi:10.1089/tmj.2019.0066
- García, L., Sanchez, G., Vazquez, E., Avalos, G., Anides, E., Nakano, M., et al. (2021). Small universal spiking neural p systems with dendritic/axonal delays and dendritic trunk/feedback. *Neural Netw.* 138, 126–139. doi:10.1016/j.neunet.2021.02.010
- Ghamisi, P., Plaza, J., Chen, Y., Li, J., and Plaza, A. J. (2017). Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geosci. Remote Sens. Mag.* 5, 8–32. doi:10.1109/mgrs.2016.2616418
- Grushin, A., Monner, D. D., Reggia, J. A., and Mishra, A. (2013). "Robust human action recognition via long short-term memory," in The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, August 04–09, 2013.
- Guo, M.-H., Lu, C.-Z., Liu, Z.-N., Cheng, M.-M., and Hu, S.-M. (2022). *Visual attention network*. arXiv preprint arXiv:2202.09741.
- Hasan, M. K., Shahjalal, M., Chowdhury, M. Z., and Jang, Y. M. (2019). Real-time healthcare data transmission for remote patient monitoring in patch-based hybrid occ/ble networks. *Sensors* 19, 1208. doi:10.3390/s19051208
- Ionescu, M., Păun, G., and Yokomori, T. (2006). Spiking neural p systems. *Fundam. Inf.* 71, 279–308.

## Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

EA, LG, and MA designed the study. TF, DH, EV, CT, and EJ developed the methods and performed experiments. JA and GS analyzed the data and wrote the paper.

## Funding

The authors would like to thank the Consejo Nacional de Ciencia y Tecnología (CONACyT) and the IPN for the financial support to realize this work.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Ji, S., Xu, W., Yang, M., and Yu, K. (2012). 3d convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 221–231. doi:10.1109/tpami.2012.59
- Liu, A.-A., Su, Y.-T., Jia, P.-P., Gao, Z., Hao, T., and Yang, Z.-X. (2014). Multiple/single-view human action recognition via part-induced multitask structural learning. *IEEE Trans. Cybern.* 45, 1194–1208. doi:10.1109/tcyb.2014.2347057
- Liu, J., Liu, H., Chen, Y., Wang, Y., and Wang, C. (2019a). Wireless sensing for human activity: A survey. *IEEE Commun. Surv. Tutorials* 22, 1629–1645. doi:10.1109/comst.2019.2934489
- Liu, Z., Liu, X., Zhang, J., and Li, K. (2019b). Opportunities and challenges of wireless human sensing for the smart iot world: A survey. *IEEE Netw.* 33, 104–110. doi:10.1109/mnet.001.1800494
- Pan, L., Păun, G., Zhang, G., and Neri, F. (2017). Spiking neural p systems with communication on request. *Int. J. Neural Syst.* 27, 1750042. doi:10.1142/s0129065717500423
- Pan, L., Zeng, X., Zhang, X., and Jiang, Y. (2012). Spiking neural p systems with weighted synapses. *Neural process. Lett.* 35, 13–27. doi:10.1007/s11063-011-9201-1
- Schuld, C., Laptev, I., and Caputo, B. (2004). Recognizing human actions: A local svm approach,” in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, 3. IEEE, 32.
- Sharma, S., Kiros, R., and Salakhutdinov, R. (2015). *Action recognition using visual attention. arXiv preprint arXiv:1511.04119.*
- Shi, Y., Zeng, W., Huang, T., and Wang, Y. (2015). “Learning deep trajectory descriptor for action recognition in videos using deep neural networks,” in 2015 IEEE international conference on multimedia and expo (ICME), Turin, Italy, June 29–July 3, 2015. (IEEE).
- Shu, N., Tang, Q., and Liu, H. (2014). “A bio-inspired approach modeling spiking neural networks of visual cortex for human action recognition,” in 2014 international joint conference on neural networks, Beijing, China, July 6–11, 2014. (IJCNN IEEE), 3450.
- Song, T., Pan, L., and Păun, G. (2014). Spiking neural p systems with rules on synapses. *Theor. Comput. Sci.* 529, 82–95. doi:10.1016/j.tcs.2014.01.001
- Veeriah, V., Zhuang, N., and Qi, G.-J. (2015). Differential recurrent neural networks for action recognition,” in Proceedings of the IEEE international conference on computer vision, Santiago, Chile, December 7–13, 2015. 4041–4049.