Latency-Based Motion Detection in Spiking Neural Networks

Mohammad Saleh Vahdatpour, Yanqing Zhang

***Abstract*—**Understanding the neural mechanisms underlying motion detection in the human visual system has long been a fascinating challenge in neuroscience and artificial intelligence. This paper presents a spiking neural network model inspired by the processing of motion information in the primate visual system, particularly focusing on the Middle Temporal (MT) area. In our study, we propose a multi-layer spiking neural network model to perform motion detection tasks, leveraging the idea that synaptic delays in neuronal communication are pivotal in motion perception. Synaptic delay, determined by factors like axon length and myelin insulation, affects the temporal order of input spikes, thereby encoding motion direction and speed. Overall, our spiking neural network model demonstrates the feasibility of capturing motion detection principles observed in the primate visual system. The combination of synaptic delays, learning mechanisms, and shared weights and delays in SMD provides a promising framework for motion perception in artificial systems, with potential applications in computer vision and robotics.

***Keywords*—**Neural networks, motion detection, signature detection, convolutional neural network.

I. Introduction

THE human visual system is a marvel of biological engineering, capable of processing and interpreting complex visual information with remarkable precision. The journey of visual information begins in the retina, where photoreceptors convert incoming light from surrounding objects into electrical signals. These signals are then relayed through a network of nerve connections to the lateral geniculate nucleus (LGN). From there, the information is further transmitted to the primary visual cortex, known as V1.

The human visual system does not stop at V1; it branches into two distinct pathways, the ventral and dorsal pathways. The dorsal pathway, often referred to as the "where" pathway, plays a critical role in guiding actions and recognizing the spatial locations of objects within a scene. Among the key regions in the dorsal pathway is the MT area, which is particularly noteworthy for its motion detection functionality. Neurons in the MT area possess spatio-temporal receptive fields, making them sensitive to a wide range of speeds and directions of moving visual stimuli.

A prominent hypothesis regarding the neuronal mechanisms underlying motion detection revolves around synaptic delays.

Synaptic delay refers to the time interval between the initiation of spike emission by a pre-synaptic neuron and the arrival of that spike at a post-synaptic neuron. This delay is primarily determined by factors such as axon length and the presence of myelin, an insulating layer around the axon that accelerates signal propagation.

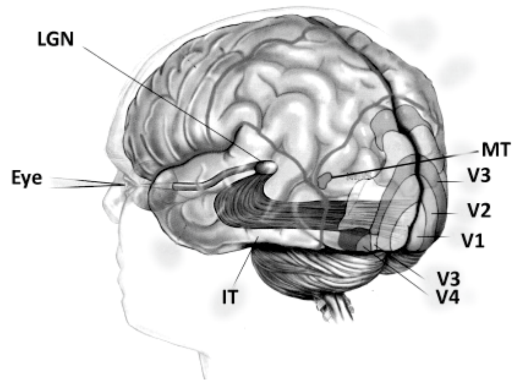


Fig. 1 Vision processes in the human brain [1]

According to this hypothesis, the order of input spikes and the time intervals between them encodes the direction and speed of motion, respectively. To design neurons that are selective to specific motion directions, they must receive input spikes from their synapses with specified time delays. This synchronization ensures that all spikes reach the neuron simultaneously, causing its membrane potential to surpass the threshold and initiate firing.

Building upon this hypothesis, our study introduces a multi-layer spiking neural network model designed to perform the task of motion detection. Neurons in this network are simulated using the Leaky Integrate and Fire (LIF) model. For the learning process, we employ a combination of Spike-Timing-Dependent Plasticity (STDP) and reinforcement learning, specifically Reward-modulated STDP. Additionally, a homeostasis mechanism is implemented to regulate neural activities.

The model is constructed in two phases. First, we create a small spiking neural network to replicate the columnar processing observed in the MT area, with each neuron becoming selective to a specific motion direction. Then, the synapses of this trained network are employed in a larger network, referred to as the Spiking Motion Detector (SMD) model, which exhibits a unique mechanism explained in detail below.

In the initial phase, we consider a 7x7 grid of neurons as the input layer. In the subsequent layer, we assign one neuron for

M. S. Vahdatpour and Y. Zhang are with the Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA (e-mail: mvahdatpour1@gsu.edu, yzhang@gsu.edu).

each desired motion direction (in our case, we consider four directions). Neurons in the first and second layers are fully connected, with random weights sampled from a uniform distribution between 0.5 and 1. To account for synaptic delays, we establish multiple synapses between neurons with different delays, selected from the set {1, 2, 3, 4}. All synapses connecting two neurons have equal initial synaptic weights.

The third layer of this model is responsible for recognizing combinations of movements detected by the second layer. Similar to the previous layers, neurons in the second and third layers are connected via multiple synapses, featuring random weights and varying delays.

For the training phase, we manually set the input stimulus for each desired motion direction to the input neurons, with a duration of 4 milliseconds. This configuration results in the second layer consisting of four neurons, while the third layer encompasses sixteen neurons to account for combinations of movements. Upon completing the learning process, we observed that each neuron in the second and third layers exhibited reinforced weights for synapses with the proper delay associated with detecting the desired motion, while the rest were suppressed.

The SMD model is subsequently constructed by incorporating the previously trained network on all possible 7x7 grids of input, considering both vertical and horizontal orientations with a stride of 1. Consequently, all neurons in the SMD that are selective to a specific motion direction or combination of two directions share the same synaptic weights (shared weights) and delays (shared delays).

The input for the SMD model consists of a sequence of frames with a temporal resolution of 1 millisecond. To adapt to different spatial resolutions in the input frames, we include a preprocessing stage that downscales high-resolution frames. This preprocessing ensures compatibility with the model's architecture.

For each input stimulus, several activity maps are generated in the second and third layers of the SMD. Each spike in these maps signifies the presence of a specific movement at a particular location within the input stimulus. The subsequent sections will delve into the experimental results of our model and its applications, particularly in online signature detection, along with a discussion of related work in the field and concluding remarks.

II. Related Works

The pursuit of understanding motion detection mechanisms in the human visual system has been a longstanding endeavor in neuroscience and artificial intelligence. To contextualize our proposed model, we delve into related work in the field, highlighting key studies and models that have contributed to our understanding of motion perception.

A. Early Motion Detection Theories

Early motion detection theories proposed the existence of specific mechanisms dedicated to perceiving motion. These theories, including the "correlation-type detectors" proposed by Adelson and Bergen [2], [3] laid the foundation for subsequent research by positing that visual motion is detected through spatial and temporal correlations in visual input. Singla [4] proposed a motion detection approach based on the frame difference method, which serves as a foundational technique in the field of motion detection

B. Synaptic Delay Hypothesis

The synaptic delay hypothesis, which forms the foundation of our model, was initially introduced by Shon et al. [5]. This hypothesis posits that the timing of spikes in neural communication, influenced by synaptic delays, encodes motion direction and speed. This idea revolutionized the understanding of motion perception by emphasizing the importance of precise timing in neural responses. Gu et al. [6] introduced the concept of spatio-temporal credit assignment with delayed feedback in deep spiking neural networks, shedding light on mechanisms for precise neural signal processing in dynamic environments.

C. Spiking Neural Networks for Motion Processing

Building on the concept of precise timing, researchers have explored the application of Spiking Neural Networks (SNNs) to model motion processing. Orchard et al. [7] presented an SNN architecture tailored for visual motion estimation, demonstrating that spiking neurons can capture the dynamics of motion perception more effectively than traditional rate-based models.

D. Learning Mechanisms in Spiking Neurons

Our model's incorporation of STDP and reinforcement learning is rooted in the work of Izhikevich [8]. His research focused on solving the distal reward problem by linking STDP with dopamine signaling. This work sheds light on the plasticity mechanisms that underlie learning in spiking neurons, enabling our model to adapt and improve over time. Fang et al. [9] introduced a novel approach that leverages deep residual learning within SNNs, showcasing advancements in the application of deep learning to motion perception.

The function that can simulate this behavior is an exponential function. Therefore, it can be stated that the exponential function serves as the foundational model for implementing the STDP learning rule. Let us consider calculating the weight changes of the synapse between neurons i_th and j_th. Let t^f_j represent the time of spiking of the pre-synaptic neuron, and t^n_i be the time of spiking of the post-synaptic neuron. The weight changes of this synapse were equal to (1):

$$\Delta w_{ij} = \sum_f \sum_n W(t_i^n - t_j^f) \tag{1}$$

where W(s) is the amount of changes using the STDP learning rule, which is described as (2):

$$W(s) = \begin{cases} A_+ \exp\left(-\frac{s}{\tau_+}\right) & if\ \ s \geq 0 \\ -A_- \exp\left(-\frac{s}{\tau_-}\right) & if\ \ s < 0 \end{cases} \tag{2}$$

where, A+, A-, -τ and +τ are the parameters, the coefficients of A+ and A- are dependent on wij and -τ and +τ are considered in about a few milliseconds [10].

E. Biological Justification and Brain Processes

Many motion detection studies lack considerations of biological justification and brain processes. For instance, experiments using deep networks for pedestrian movement detection report an accuracy of about 68%, highlighting the challenges of observational learning in deep networks [11]. Neural network studies show that tapping of neurons is crucial for recognizing different movements and speeds, emphasizing the importance of biologically inspired learning [7], [17].

F. Models Based on Hidden Markov Model, Dynamic Scheduling, and Support Vector Machine

Various models, such as those based on hidden Markov models [12], dynamic scheduling [13], and feature identification using support vector machines [14], have been proposed. Notably, models based on dynamic scheduling have shown improved accuracy compared to hidden Markov models [13]. Support vector machine algorithms, particularly those extracting new features with lower dimensions, have also demonstrated efficacy in motion detection [14].

G. Online Signature Detection

Our study applies motion detection principles to the domain of online signature detection. While motion detection and signature verification may seem unrelated, they share common ground in the analysis of dynamic patterns. In this context, the SUSIG dataset [15] has become a pivotal resource for benchmarking signature detection algorithms, offering a diverse set of genuine and forgery signatures.

Our proposed model synthesizes insights from these diverse areas of research. By combining the synaptic delay hypothesis, SNNs, and learning mechanisms, we aim to create a model that not only advances our understanding of motion perception but also demonstrates practical applicability in real-world tasks such as online signature verification. The ensuing sections of this paper detail our model's architecture, experiments, and results, showcasing its potential for motion-based pattern recognition

III. A Latency-Based Motion Detection Model

Our proposed method is rooted in the hypothesis that synaptic delays are a pivotal factor in the neural encoding of motion direction and speed. We present a comprehensive multi-layer spiking neural network model, designed to simulate the intricate processes involved in motion detection within the primate visual system, with a particular focus on the MT region.

A. MT Columnar Processing (Phase 1)

In the first phase of our model, we aim to replicate the columnar processing observed in the MT area, where neurons become selectively tuned to specific motion directions. The foundation of this phase lies in a 7x7 grid of neurons designated as the input layer. Subsequently, we create a layer with one neuron allocated for each desired motion direction, considering four distinct directions in our study.

Neurons in the input and second layers are fully connected, and their connectivity is characterized by synaptic weights that are randomly initialized from a uniform distribution within the range of 0.5 to 1. To introduce the critical element of synaptic delays, we employ multiple synapses connecting neurons with different delay values selected from the set $\{1, 2, 3, 4\}$. Notably, all synapses between any two neurons in this layer initially possess equal synaptic weights.

The third layer of our model takes on the responsibility of recognizing combinations of movements detected by the second layer. Like the preceding layers, neurons in the second and third layers are interconnected via multiple synapses, each featuring random weights and distinct delays.

During the learning phase, we manually set the input stimulus for each desired motion direction. Each stimulus is characterized by a fixed duration of 4 milliseconds. As a result of this configuration, the second layer consists of precisely four neurons, whereas the third layer encompasses sixteen neurons, each dedicated to detecting specific combinations of movements. Upon the completion of the learning process, we observe that the synaptic weights associated with the proper delays for detecting the desired motion are reinforced, while those linked to other delays are suppressed.

B. Spiking Motion Detector (SMD) Model (Phase 2)

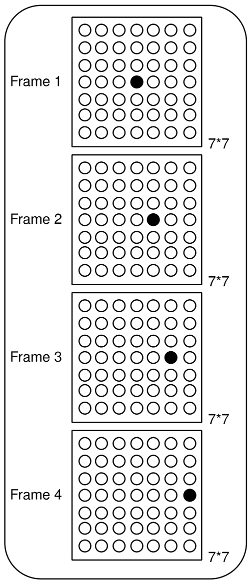
In the second phase, we construct the SMD model, which leverages the previously trained network from the first phase. This model extends its applicability to all possible 7x7 grids of input, encompassing both vertical and horizontal orientations with a stride of 1. Importantly, all neurons within the SMD model that are selective to specific motion directions or combinations of two directions share identical synaptic weights (shared weights) and delays (shared delays).

The input for the SMD model consists of a sequence of frames, each offering a temporal resolution of 1 millisecond. To accommodate input frames with diverse spatial resolutions, we incorporate a preprocessing stage that down-scales high-resolution frames. This preprocessing ensures the congruence of input data with the architectural specifications of the model.

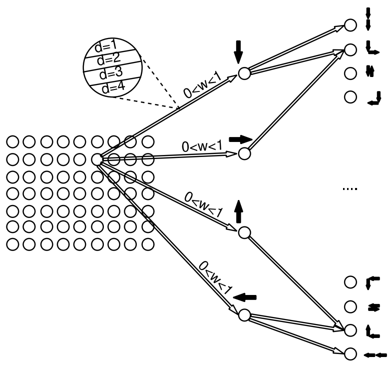
For each input stimulus, the SMD model generates multiple activity maps in the second and third layers. Each spike within these maps serves as an indicator of the presence of a particular movement at a designated location within the input stimulus. These activity maps collectively capture the intricate patterns of motion within the visual input.

IV. Experimental Results

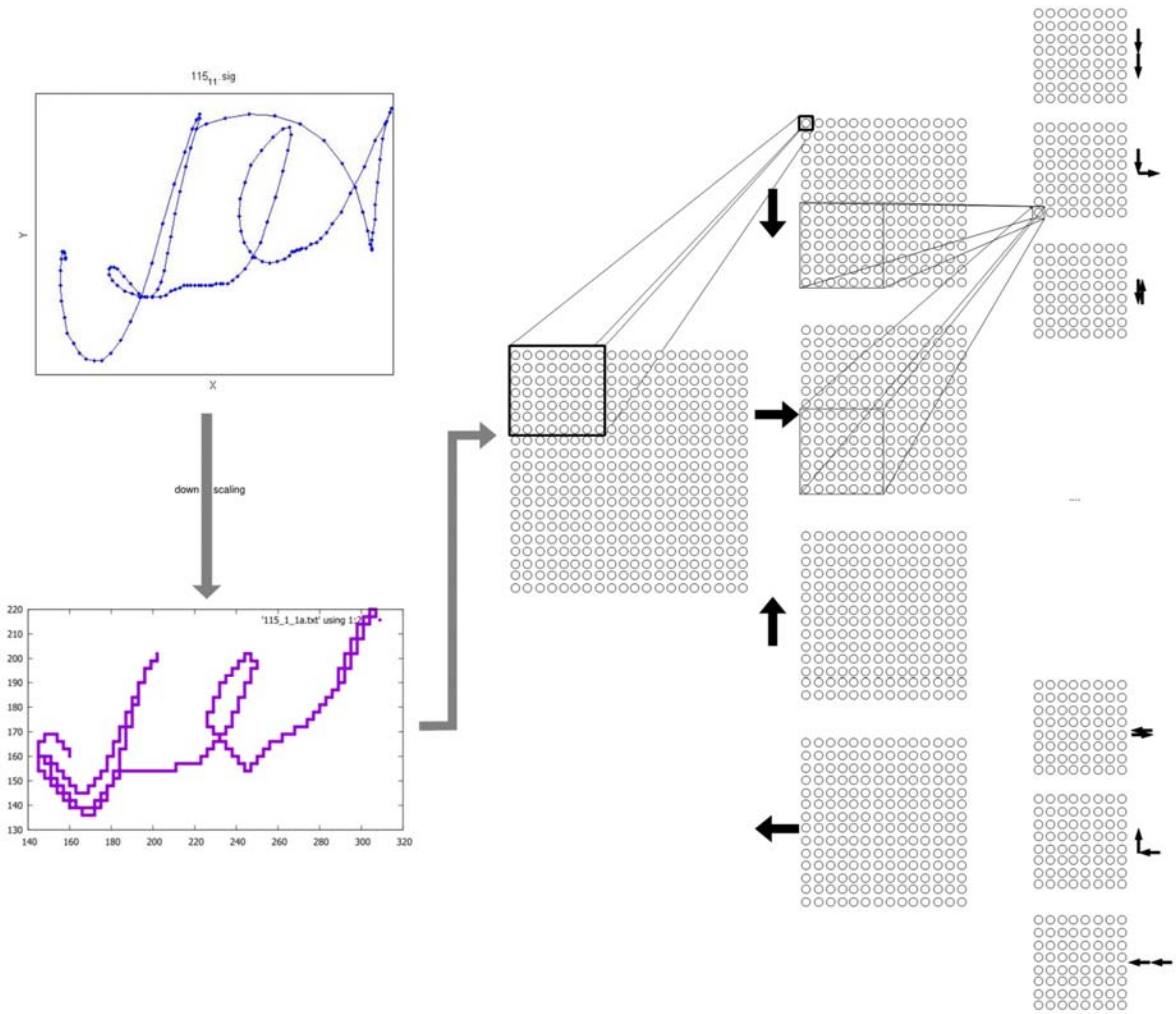
To rigorously evaluate the performance of our SMD model, we conducted extensive experiments, with a specific focus on the task of online signature detection. This evaluation was performed using the prestigious SUSIG dataset [8], which includes 940 genuine signatures from 94 individuals, along with 940 forgery signatures, all recorded at high resolution.



(a)



(b)



(c)

Fig. 2 (a) Input sample (right direction), (b) Small spiking neural network for training direction sensitive neurons, and (c) SMD; model is constructed by considering network shows in part b on all possible 7×7 grids of input

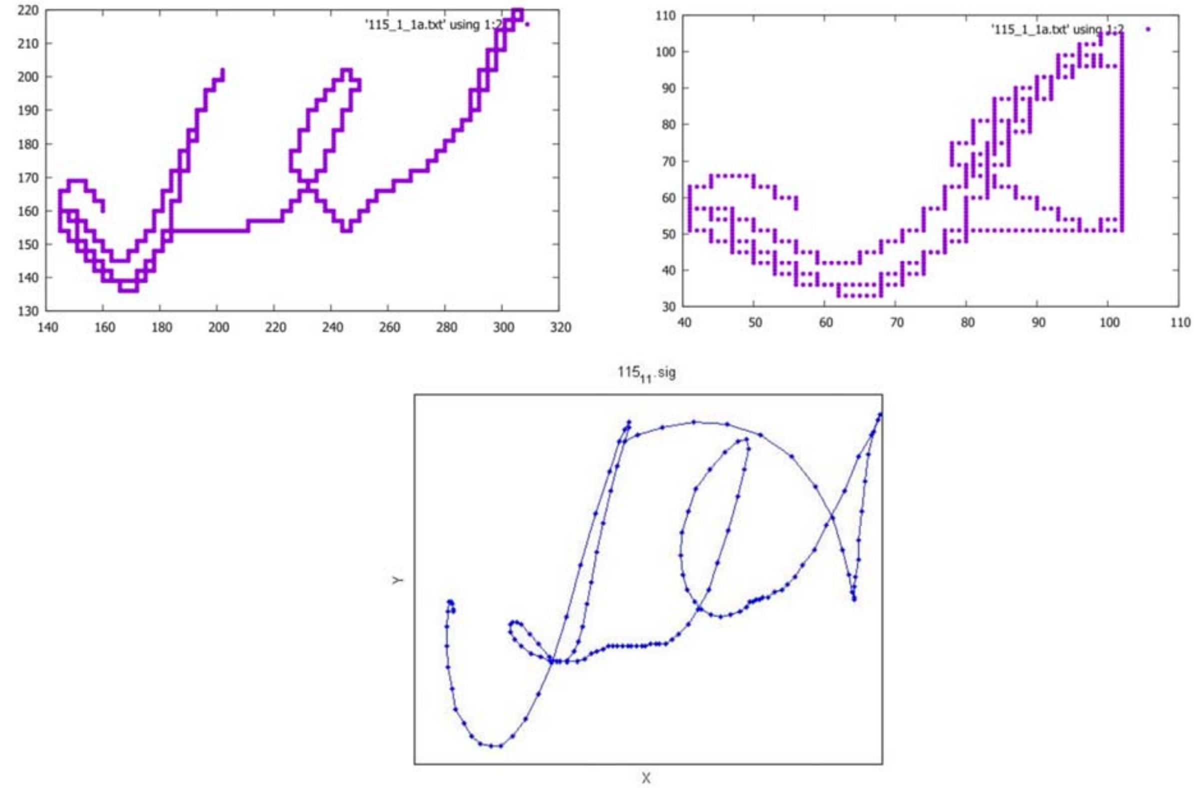


Fig. 3 Sample signature in database and different scales

A. Data Preprocessing and Feature Extraction

In preparation for the experiments, we undertook data preprocessing steps to align the dataset with our model's architecture. This included down-scaling all signatures by a uniform factor of 300. This transformation effectively converted the continuous pen motions inherent in the signatures into the trained motions employed within our model. Various tests showed that if we consider one movement for every 300 points between the points, while compressing the input data, the overall structure of the signature is also preserved. Certainly, reducing this number brings the input signature closer to the original structure, but at the same time, it also increases the processing time and complexity. On the other hand, scaling to 1000 compresses the data and increases the processing speed, but the overall structure of the signature is also completely lost. For the signatures in the data, we convert the movements and directions in the signatures into four identifiable directions in the model and give the output of the model to classification algorithms. One of the signatures in the data along with the scales of 300 and 1000 is shown in Fig. 3.

Crucially, we constructed feature vectors to represent each signature in a manner conducive to motion detection. These feature vectors encapsulate the number of spikes emitted by neurons in each map of the third layer. Importantly, they not only encode the presence or absence of specific motions but also retain temporal information, capturing the intricate dynamics of signature movements. The feature vector, X, for each signature is calculated (3):

$$X_i = \sum_{j=1}^{N} S_{ij} \quad for\ i = 1,2, \dots, M \tag{3}$$

where N is the number of neurons in the third layer, M is the number of signatures, and S_{ij} is the spike count of the j-th neuron in response to the i-th signature.

B. Detection and Classification

For the final detection task, we employed a diverse set of widely recognized classifiers, including K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), Simple Logistic, and Naive Bayes. These classifiers played a pivotal role in determining the authenticity of signatures based on the feature vectors extracted from the activity maps of the SMD model.

The classification decision for a given signature i is determined by the classifier as (4):

$$Class(i) = \begin{cases} Forgery & if\ P(i) < \theta \\ Genuine & otherwise \end{cases} \tag{4}$$

where P(i) is the probability assigned by the classifier for signature i being genuine, and θ is a classification threshold.

The evaluation of our model's performance was conducted using a robust cross-validation technique, ensuring the reliability and generalizability of the results. Through this process, we assessed the model's ability to distinguish between genuine and forgery signatures accurately.

The experimental results provide compelling evidence of the SMD model's proficiency in online signature detection. Table I illustrates the performance metrics achieved by our model.

TABLE I
ACCURACY OF THE PROPOSED MODEL TO DETECT SUSIG SIGNATURES

Classification algorithms	Number of class	Genius sig		
		AUC	Accuracy	
			Classification	Forgery detection
Random Forest	40	0.999	100	99.3
Random Forest	94	0.996	85.2	98.4
KNN	94	0.981	81.9	84.9
Naïve Bayes	94	0.987	69.7	66.2
SVM	94	0.831	73.4	67.8
Simple Logistic	94	1	100	97.2

As indicated in Table I, our model successfully detected a remarkable 98.4% of forgery signatures, demonstrating its exceptional accuracy and effectiveness in authenticating signatures. This high level of accuracy underscores the practical applicability of our model in real-world security and authentication scenarios.

V. CONCLUSIONS

In conclusion, while our designed model exhibits high proficiency in detecting movements, it is not without weaknesses and deficiencies. One potential avenue for improvement lies in strengthening the training phase. Currently, the model identifies only four main directions, a limitation that could be addressed by extending recognition to eight directions through mechanism adjustments. Research indicates that the human brain recognizes eight directions and approximates other directions from these primary orientations [16]. However, it is crucial to note that generalizing the model to more directions increases complexity and training time. Expanding the model to recognize eight directions leads to an expansion in the layer of combined movements, resulting in the recognition of 64 combined directions. Furthermore, the current use of single-shot, aggregation, and leaky neurons limits the model's ability to identify directions between the recognized primary orientations. Future work involves exploring alternative neuron models and structural changes to incorporate mechanisms for identifying intermediary directions, similar to the human vision system.

VI. FUTURE WORKS

Moving forward, our exploration of latency-based motion detection in SNNs suggests several avenues for improvement. Key priorities include fine-tuning synaptic delay parameters to enhance precision, scaling up the model for real-world testing, introducing dynamic adaptation mechanisms, and aligning the model more closely with the human visual system. Additionally, we plan to assess transfer learning capabilities and explore hardware implementations. These efforts aim to refine the model's theoretical foundations, ensuring its adaptability and practical applicability across diverse domains such as artificial intelligence, computer vision, and robotics.

REFERENCES

[1] Frisby, John P., and James V. Stone. “Seeing: The computational approach to biological vision.” The MIT Press, 2010.

[2] Borst, Alexander, and Martin Egelhaaf. “Principles of visual motion detection.” Trends in neurosciences 12.8 (1989): 297-306.

[3] Beauchemin, Steven S., and John L. Barron. “The computation of optical flow.” ACM computing surveys (CSUR) 27.3 (1995): 433-466.

[4] Singla, Nishu. “Motion detection based on frame difference method.” International Journal of Information & Computation Technology 4.15 (2014): 1559-1565.

[5] Shon, Aaron P., Rajesh PN Rao, and Terrence J. Sejnowski. “Motion detection and prediction through spike-timing dependent plasticity.” Network: Computation in Neural Systems 15.3 (2004): 179-198.

[6] Gu, Pengjie, et al. “STCA: Spatio-temporal credit assignment with delayed feedback in deep spiking neural networks.” IJCAI. Vol. 15. 2019.

[7] Orchard, Garrick, et al. “A spiking neural network architecture for visual motion estimation.” Biomedical Circuits and Systems Conference (BioCAS), 2013 IEEE. IEEE, 2013.

[8] Izhikevich, Eugene M. “Solving the distal reward problem through linkage of STDP and dopamine signaling.” Cerebral cortex 17.10 (2007): 2443-2452.

[9] Fang, Wei, et al. “Deep residual learning in spiking neural networks.” Advances in Neural Information Processing Systems 34 (2021): 21056-21069.

[10] Masquelier, Timothée, and Simon J. Thorpe. “Unsupervised learning of visual features through spike timing dependent plasticity.” PLoS computational biology 3, no. 2 (2007): e31

[11] Angelova, Anelia, Alex Krizhevsky, Vincent Vanhoucke, Abhijit S. Ogale, and Dave Ferguson. “Real-Time Pedestrian Detection with Deep Network Cascades.” In BMVC, vol. 2, p. 4. 2015.

[12] Pylvänäinen, Timo. “Accelerometer based gesture recognition using continuous HMMs.” Pattern Recognition and Image Analysis (2005): 413-430.

[13] Liu, Jiayang, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. “uWave: Accelerometer-based personalized gesture recognition and its applications. ” Pervasive and Mobile Computing 5, no. 6 (2009): 657-675.

[14] Wu, Jiahui, Gang Pan, Daqing Zhang, Guande Qi, and Shijian Li. “Gesture recognition with a 3-d accelerometer.” In International Conference on Ubiquitous Intelligence and Computing, pp. 25-38. Springer Berlin Heidelberg, 2009.

[15] Kholmatov, Alisher and Yanıkoglu, Berrin (2009) “SUSIG: an on-line signature database, associated protocols and benchmark results.” Pattern Analysis & Applications, 12 (3). pp. 227-236. ISSN 1433-7541.

[16] Vass, Lindsay K., and Russell A. Epstein. “Abstract representations of location and facing direction in the human brain.” Journal of Neuroscience 33, no. 14 (2013): 6133-6142.

[17] Vahdatpour, Mohammad Saleh. "Addressing the Knapsack Challenge through Cultural Algorithm Optimization." CS & IT Conference Proceedings. Vol. 13. No. 19. CS & IT Conference Proceedings, 2023.