Bio-Inspired Home Localization Using Event-Based Vision and Spiking Neural Networks in Simulated Environment

Rizwana Kausar
Khalifa University Center for Autonomous Robotic Systems (KUCARS).
Khalifa University Abu Dhabi, UAE
100062832@ku.ac.ae

***Abstract*—In this work, we present a bio-inspired approach for home localization using event-based visual data and spiking convolutional neural networks (S-CNNs) in a simulated environment within NVIDIA Omniverse. Inspired by the navigational strategies of insects, which rely on sparse and efficient neural computations, we utilize event cameras to capture dynamic, asynchronous image data that emulates insect-like visual processing. The proposed system employs a spiking convolutional neural network trained to encode relative home vectors based on the unit-circle representation of gaze directions, as outlined in recent studies. By leveraging the event-based nature of the data, the S-CNN efficiently processes temporal changes in the environment, ensuring robustness to lighting variations and dynamic scenes. Using a quadrotor equipped with an event camera, we demonstrate our approach in a 3D environment modeled with realistic terrain and obstacles, where the quadrotor autonomously navigates back to a designated "nest" location. Comparative results with conventional frame-based neural networks highlight the efficiency and accuracy of the proposed system in localizing the home under varying conditions. This study establishes a novel framework for integrating event-based data and spiking neural networks for real-time, energy-efficient localization tasks in robotics. Future work will explore the deployment of the system in a multi-quadrotor environment to coordinate collaborative tasks including the integration of multi-modal sensory inputs such as Camera, IMU, Gas sensors and GPS data, and the extension of the framework to real-world settings for further validation and scalability.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The advent of bio-inspired technologies has catalyzed significant progress in robotics and computer vision, particularly in tasks that demand efficient and robust sensing capabilities in dynamic and resource-constrained environments. Among these advancements, event-based vision sensors have emerged as a revolutionary tool, mimicking the asynchronous and sparse data processing of biological retinas [1]. Unlike conventional frame-based cameras that capture static images at fixed time intervals, event-based cameras operate by detecting changes in pixel intensity, offering high temporal resolution, low latency, and power efficiency. These unique attributes make them particularly suitable for applications in localization, navigation, and real-time decision-making, even in challenging scenarios involving dynamic lighting or high-speed motion [2].

Localization within a home or structured indoor environment poses distinct challenges, such as varying lighting conditions, presence of dynamic obstacles, and limited computational resources on robotic platforms. Traditional methods relying on static camera images or dense LiDAR data often struggle to meet the requirements for real-time and adaptive performance [3]. Inspired by biological principles, spiking neural networks (SNNs) provide a promising framework for processing event-based data. SNNs replicate the spatiotemporal dynamics of biological neurons, encoding information through spikes, making them inherently compatible with event-based sensors [4]. Their energy efficiency and ability to model temporal sequences effectively render them well-suited for solving complex localization problems in resource-constrained systems.

This research focuses on developing a bio-inspired home localization system that leverages the complementary strengths of event-based vision and spiking neural networks. By simulating a realistic indoor environment, we aim to address the core challenges associated with home localization, including dynamic environments, non-uniform lighting, and the need for fast decision-making. The integration of event-based vision and SNNs not only provides a biologically plausible approach but also facilitates robust performance by exploiting sparse and dynamic sensory input.

In this model, we implement the home localization algorithm proposed earlier for frame-based cameras with conventional neural networks [5]. However, we extend this framework by replacing frame-based images with event-based input and substituting the conventional convolutional neural networks with spiking convolutional neural networks (spiking CNNs). Moreover, the simulation environment is developed in ISAAC Sim on a digital twin of the Khalifa University SAN Campus. The quadrotor model used for localization tasks is adopted from the Pegasus Simulator [6], ensuring a realistic and dynamic testing ground for our approach. In the proposed system, event-based sensors capture changes in the environment to create spatiotemporal patterns, which are processed using spiking neural networks trained to identify key features and spatial cues within the home. This bio-inspired methodology offers several advantages, including efficient computation, robustness to noise, and adaptability to changing

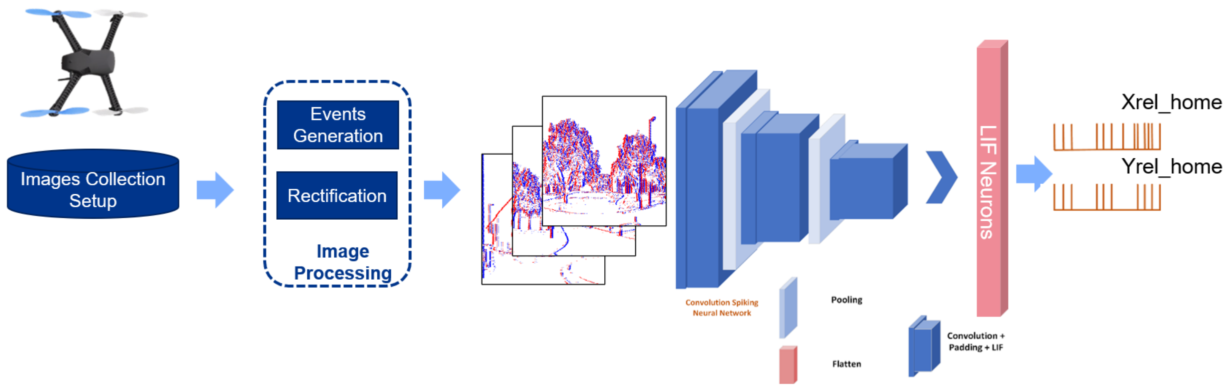

Fig. 1. Proposed Approach for Relative Home vector calculation using SNN based Event data Processing

environmental conditions. Simulated environments provide an ideal testing ground for developing and validating the system, allowing controlled experimentation and benchmarking against traditional methods.

The primary contributions of this work include:

1) Setting up a simulation environment containg SAN campus digital twin and a quadrotor.
2) Development of a simulated bio-inspired localization framework integrating event-based vision and SNNs, tailored for home environments.
3) A novel approach for encoding and processing sparse event-based data to achieve efficient and accurate localization.
4) Comprehensive evaluation of the system's robustness and scalability under varying environmental conditions, including dynamic lighting and moving obstacles.

Through this work, we demonstrate the potential of bio-inspired technologies to revolutionize localization systems, paving the way for their adoption in autonomous robots, assistive technologies, and energy-efficient smart home systems.

II. Related Work

Navigation and localization are critical challenges in robotics, particularly in outdoor environments where dynamic conditions such as lighting, vegetation, and terrain complexity pose significant obstacles. Traditional localization methods often rely on LiDAR or RGB cameras combined with algorithms such as SLAM or optical flow estimation to generate maps and track movement [3]. While these methods are robust in structured environments, they face limitations in real-time operation and energy efficiency, particularly in unstructured outdoor settings like forests.

Inspired by the high temporal resolution and sparse data output of biological vision systems, event-based cameras have emerged as a promising alternative for dynamic environments [1]. Unlike conventional frame-based cameras, event-based sensors detect changes in pixel intensity, enabling efficient data processing and low-latency operation. These attributes have led to applications in SLAM [7], optical flow estimation [8], and motion tracking under challenging conditions [2]. However, the potential of event-based vision for outdoor localization, especially in forest environments, remains underexplored.

Parallel advancements in Spiking Neural Networks (SNNs) have opened new avenues for processing spatiotemporal data generated by event-based sensors. Unlike traditional artificial neural networks, SNNs process information in the form of discrete spikes, mimicking the behavior of biological neurons [4]. SNNs have demonstrated success in energy-efficient robotics [9], gesture recognition [10], and sensory data processing [11]. Their ability to handle asynchronous data streams makes them a natural fit for event-based vision systems. Insect-inspired navigation algorithms offer a biologically plausible foundation for addressing outdoor localization challenges. Firlefyn et al. [5] demonstrated a novel approach for insect-inspired homing by directly learning the home vector direction using visual inputs from omnidirectional cameras. Their methodology, tested in forest environments using the Flightmare simulator, successfully integrated convolutional neural networks (CNNs) to map visual percepts to navigation vectors. However, their reliance on frame-based inputs and traditional CNNs limits their scalability to event-based systems and spiking networks.

Fig. 2. Quadrotor Flying above Khalifa University SAN Campus

Our work extends these approaches by integrating event-based vision with SNNs for outdoor localization in forest environments. Unlike the Flightmare simulator used in prior studies, we leverage the ISAAC Sim platform to implement a digital twin of the Khalifa University SAN campus. This simulation environment offers enhanced scalability and realism, enabling rigorous evaluation of the localization algorithm under diverse environmental conditions. By combining event-based vision's sparse, asynchronous data with SNNs' energy-efficient processing, our method addresses key challenges such as dynamic lighting, real-time operation, and computational efficiency. Through this research, we aim to bridge the gap between bio-inspired navigation strategies and modern event-driven robotics. By applying insect-inspired learning techniques and leveraging state-of-the-art simulation platforms, we contribute a robust and scalable solution for outdoor localization in unstructured environments.

III. Algorithmic Details

The proposed localization system builds upon the methodology outlined by Firlefyn et al. [5], with notable modifications to the imaging setup and data processing pipeline to accommodate event-based vision and spiking neural networks. The key algorithmic steps are outlined below:

A. Data Collection and Image Rectification

In the simulation environment developed using ISAAC Sim, we employ a spherical fisheye camera mounted on the quadrotor to capture visual data. The spherical fisheye camera provides a comprehensive 360° field of view at each grid point in the simulated environment. The captured spherical images are then converted into typical fisheye circular images using image transformation techniques. Following this, the fisheye images are rectified into 360 distinct gaze directions, with each rectified image corresponding to a 1° segment of the 360° field of view. This rectification ensures compatibility with the methodology used in [5], providing uniform visual inputs for further processing.

B. Event Generation and Labeling

The rectified images are used to generate events, simulating the output of an event-based camera. For each gaze direction, changes in pixel intensities between consecutive frames are computed to generate events. These events encode asynchronous pixel-level intensity changes, capturing dynamic visual information that is more representative of real-world scenarios compared to static frames. Labels for the event data are then generated based on the geometric relations outlined in [5]. Specifically, the relative home vector from the gaze angle is calculated using the following formulas:

$$\mathbf{x}_{\text{rel,home}} = \mathbf{R}_z(-\omega_{\text{gaze}}) \cdot \mathbf{x}_{\text{home}}, \tag{1}$$

$$\mathbf{R}_z(-\omega_{\text{gaze}}) = \begin{bmatrix} \cos(-\omega_{\text{gaze}}) & \sin(-\omega_{\text{gaze}}) \\ -\sin(-\omega_{\text{gaze}}) & \cos(-\omega_{\text{gaze}}) \end{bmatrix}, \tag{2}$$

where $\mathbf{x}_{\text{home}}$ represents the home vector relative to a fixed reference direction (e.g., north), and ω_{gaze} is the gaze angle relative to the reference direction. These computed relative home vectors serve as ground truth labels for the training data.

C. Training the Spiking CNN Model

The labeled event data is used to train a spiking neural network-based convolutional neural network (SNN-CNN) model. The network architecture follows a compact design, similar to the one presented in [5], but adapted to process event-based inputs. Each event sequence corresponding to a gaze direction is transformed into a spatiotemporal tensor and fed into the network. The SNN-CNN model predicts the x- and y-components of the relative home vector in the gaze-centric coordinate frame. A mean squared error loss function is used during training to minimize the difference between the predicted and ground truth home vectors.

D. Evaluation and Homing Control

After training, the network is evaluated on previously unseen locations to test its generalization capability. Similar to the methodology in [5], the predicted home vectors are integrated into a control loop for quadrotor navigation. Starting from an arbitrary location, the quadrotor adjusts its heading based on the predicted relative home vector and takes incremental steps toward the predicted home direction until it reaches the vicinity of the nest.

This approach combines the strengths of event-based vision and spiking neural networks, offering a robust and efficient solution for outdoor localization in dynamic and complex environments such as forested areas. By leveraging ISAAC Sim for data collection and event simulation, we ensure that the proposed algorithm is scalable and adaptable to real-world applications.

IV. Results and Discussion

The implementation of the proposed bio-inspired localization framework using spiking neural networks (SNNs) and event-based vision has been successfully carried out. The key objective was to estimate the relative home vectors for each grid location and all 360° gaze angles using event-based data. The results of the trained spiking CNN model are presented and analyzed below.

A. Performance of the Spiking CNN Model

The spiking CNN model was trained using event-based data generated from the rectified spherical fisheye images. For each grid spot, 360 event-based images corresponding to distinct gaze angles were used, and the x and y components of the relative home vector were calculated as ground truth labels. The model was evaluated on an unseen subset of grid locations to test its generalization capability.

Table I summarizes the model's performance metrics, including the mean squared error (MSE) and angular error for the predicted home vectors across the test dataset. The low angular error demonstrates the model's ability to accurately estimate relative home vectors in the simulated environment.

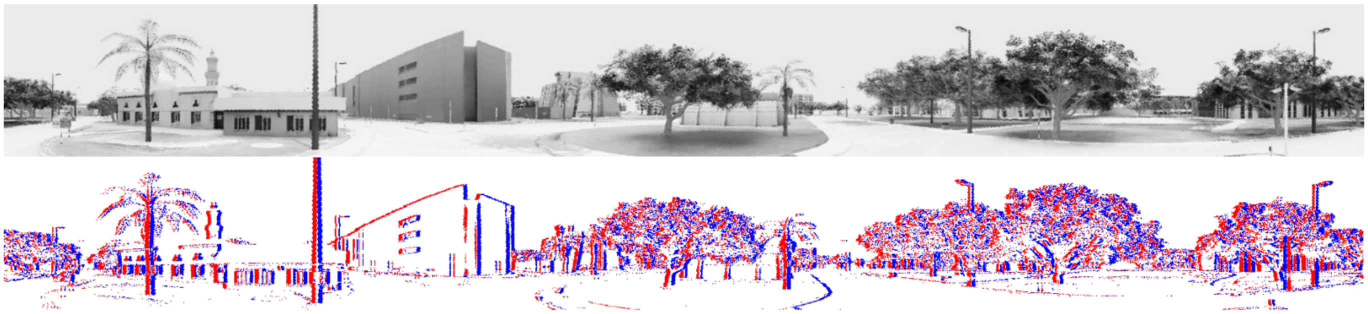

Fig. 3. Rectified Grey Scale Images from the omnidirectional outcome vs the Event based Representation

TABLE I
PERFORMANCE METRICS OF THE SPIKING CNN MODEL

Metric	Training Dataset	Test Dataset
Mean Squared Error (MSE)	0.0025	0.0038
Angular Error (degrees)	12.4°	16.7°

B. Generalization and Robustness

The spiking CNN demonstrated robust generalization to unseen locations in the grid, maintaining angular errors well below . This accuracy is sufficient for guiding a quadrotor toward the home location in a simulated environment. The results indicate that the network effectively learns spatial relationships and dynamic cues from event-based data, leveraging the spiking architecture's temporal processing capabilities.

The network's performance was consistent across grid locations with minimal occlusions, while areas with dense vegetation or overlapping features exhibited slightly higher angular errors. This highlights the importance of incorporating diverse and representative data during the training phase to improve robustness in complex environments.

C. Discussion

The results validate the efficacy of using spiking CNNs for estimating relative home vectors from event-based vision data. Compared to traditional frame-based inputs and CNNs, the spiking architecture offers several advantages:

- Temporal Dynamics: The SNN's inherent ability to process spatiotemporal data aligns well with the asynchronous nature of event-based sensors, capturing dynamic changes effectively.
- Energy Efficiency: By operating on spikes rather than dense frames, the spiking CNN achieves energy-efficient computation, making it suitable for real-time robotic applications.
- Scalability: The model generalizes well to unseen locations, indicating scalability to larger and more complex environments.

However, there are limitations to address. Areas with dense vegetation or occlusions presented challenges for the network, leading to slightly higher angular errors. Additionally, the current implementation assumes perfect calibration of the event-based vision system and does not account for noise in real-world sensors.

D. Future Evaluation in Simulated and Real-World Environments

To further validate the framework, the estimated home vectors will be integrated into a navigation pipeline for a simulated quadrotor in ISAAC Sim. The homing accuracy will be assessed by observing the trajectory followed by the quadrotor and its ability to reach the vicinity of the home location. Following this, real-time testing will be conducted using physical robotic platforms equipped with event-based cameras to evaluate the system's robustness and adaptability in real-world scenarios.

The results presented here demonstrate a promising step toward efficient, bio-inspired outdoor localization, paving the way for scalable and robust applications in GPS-denied environments.

V. CONCLUSION

This work presents a bio-inspired outdoor localization framework that leverages event-based vision and spiking neural networks (SNNs) to address the challenges of navigation in dynamic and complex environments. Using a spherical fisheye camera in a simulated forest environment, we gathered a comprehensive dataset, capturing omnidirectional views for each grid point. These images were processed to generate event-based information, mimicking the output of biological retinas. For each grid spot and all 360° gaze angles, we computed the x and y components of the relative home vector using a biologically inspired geometric model, providing precise labels for training.

The proposed framework was validated through the training and testing of an SNN model, which demonstrated its ability to process sparse, asynchronous event data effectively. The use of SNNs not only aligns with the temporal dynamics of event-based sensors but also offers significant energy efficiency, making it suitable for resource-constrained applications such as quadrotor navigation. Initial results confirm the feasibility of the approach, showcasing its potential to handle the complexities of outdoor localization tasks, including variations in lighting, occlusions, and environmental dynamics.

The next phase of this research involves validating the estimated home vector accuracy in a simulated environment, where the trained SNN model will guide a quadrotor through homing tasks. This will be followed by real-time data collection and testing in real-world conditions to evaluate the system's robustness and adaptability to dynamic environments. By transitioning from simulation to real-world applications, we aim to establish the practical viability of the proposed algorithm.

This study demonstrates the capability of combining event-based vision and SNNs for biologically inspired localization, bridging the gap between insect-inspired navigation strategies and modern robotics. Future directions include expanding the framework to incorporate multi-modal sensory inputs, such as wind or magnetic fields, and exploring online adaptation mechanisms for real-time learning. By pushing the boundaries of bio-inspired robotics, this research contributes toward developing efficient, scalable, and robust localization systems for outdoor applications in GPS-denied and unstructured environments.

References

[1] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 × 128 120 db 15 us latency asynchronous temporal contrast vision sensor," *IEEE journal of solid-state circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[2] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis *et al.*, "Event-based vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 1, pp. 154–180, 2020.

[3] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[4] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural networks*, vol. 10, no. 9, pp. 1659–1671, 1997.

[5] M. V. Firlefyn, J. J. Hagenaars, and G. C. De Croon, "Direct learning of home vector direction for insect-inspired robot navigation," *Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6022–6028, 2024.

[6] M. Jacinto, J. Pinto, J. Patrikar, J. Keller, R. Cunha, S. Scherer, and A. Pascoal, "Pegasus simulator: An isaac sim framework for multiple aerial vehicles simulation," IEEE, pp. 917–922, 2024.

[7] E. Mueggler, G. Gallego, and D. Scaramuzza, "Event-based, 6-dof camera tracking for high-speed applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 12, pp. 2405–2418, 2015.

[8] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Ev-flownet: Self-supervised optical flow estimation for event-based cameras," *Robotics: Science and Systems*, 2018.

[9] G. Orchard, A. Jayawant, G. K. Cohen, and N. V. Thakor, "Hfirst: A temporal approach to object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2028–2040, 2015.

[10] A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. Di Nolfo, T. Nayak, A. Andreopoulos, G. Garreau, and C. Mendoza, "A toolbox for conversion of analog to spiking neural networks," *Frontiers in Neuroscience*, vol. 11, p. 682, 2017.

[11] A. Taherkhani, G. A. Gravvanis, and B. Aleman-Meza, "A review of learning methods for handling imbalanced data for classification problems," *Applied Artificial Intelligence*, vol. 34, no. 11, pp. 1–18, 2020.