3iA Côte d'Azur
Institut interdisciplinaire d'intelligence artificielle

Post-Doctoral research topic

Benoît Miramond, benoit.miramond@univ-cotedazur.fr[1]

[1]Université Côte d'Azur / LEAT / CNRS , Sophia-Antipolis, France

Research axis of the 3IA: AI for Computional Biology and **Bio-inspired AI**

The laboratory:
Benoît Miramond is Full Professor in Electrical Engineering at **LEAT** laboratory from University Côte d'Azur (UCA). He leads the eBRAIN research group and develops a interdisciplinary research activity on embedded Bio-inspiRed AI and Neuromorphic architectures, especially based on SNNs. LEAT is a mixt research unit (UMR 7248) from UCA and CNRS.

Abstract

Optical flow estimation is a key task in computer vision, particularly critical for autonomous navigation where accurate motion perception is essential. It can be integrated with various other vision tasks—such as semantic segmentation, object detection, tracking, and motion analysis—to enhance temporal resolution and overall scene understanding.

Conventional approaches primarily rely on frame-based cameras. However, these cameras, typically operating at 20–30 frames per second, suffer from several limitations: motion blur and glare reduce the quality of information within frames, while the low temporal resolution leads to significant data loss between frames. Although increasing the frame rate could alleviate some of these issues, it comes at the cost of higher computational load and energy consumption, due to the processing of largely redundant pixel data. Moreover, this does not resolve intrinsic limitations such as narrow dynamic range.

Event cameras offer an alternative by capturing asynchronous changes in brightness at the pixel level, enabling high temporal resolution and low-latency sensing. Nonetheless, most event-based methods depend on supervised learning using ground truth aligned with frame-based sensors, which inherently limits their temporal precision. Meanwhile, self-supervised methods—such as those based on contrast maximization—remain far less accurate than their supervised counterparts and often demand significant computational and energy resources.

To address these challenges, we propose a novel hybrid vision architecture that leverages both frame and event data to estimate optical flow at high temporal resolution. This approach aims to combine the complementary strengths of both sensing modalities, enabling robust and efficient motion estimation with improved accuracy and reduced computational cost.

Semantic segmentation is crucial for autonomous navigation systems, yet current frame-based approaches face fundamental limitations. Traditional cameras operating at 20-30 fps experience information loss through motion blur and glare within frames, while their low temporal resolution results in missed data between frames. Although increasing frame rates could mitigate temporal resolution issues, this approach leads to higher computational costs and energy consumption by processing largely redundant pixel data, without addressing fundamental dynamic range constraints. To address these challenges, we propose a research project to develop a novel hybrid vision architecture combining conventional frame-based cameras with event-based sensors capable of microsecond-scale temporal resolution. The targeted system integrates two key components: (1) an energy-efficient Spiking Neural Network (SNN) for frame-based semantic segmentation, and (2) an adaptive mechanism using event-based optical flow to interpolate semantic information between frames. Our approach will be evaluated on existing event-based dataset (DSEC, MVSEC...) on semantic segmentation by compensating for information loss within frames thanks to the continuous steam of events. Moreover, integrating optical flow for both

semantic segmentation and motion estimation eliminates the need for separate neural networks, reducing computational complexity and improving the overall efficiency of autonomous navigation.

Keywords: bio-inspired computing, spiking neural networks, event-based sensors, supervised and unsupervised learning, image processing, neuromorphic systems, low-power processing, embedded applications

Application:
Apply by sending an email directly to the supervisor. The application will include:

- Letter of recommendation of the supervisor indicated above.
- Curriculum vitæ.
- Motivation Letter.
- At least, one letter of recommendation.

1 Context

The "neuromorphic" event-based approach to vision and image sensing is recently gaining substantial attention as it proposes solutions to the problems encountered with conventional technology of image processing. The output of such a sensor is a time-continuous stream of pixels, delivered at unprecedented temporal resolution, containing zero redundancy and encoding orders of magnitude higher dynamic range than conventional image sensors. However, due to the lack of alternatives so far, the event-based, asynchronous output of these sensors have been processed using conventional computing devices such as CPUs and GPUs. This way of processing is obviously non-ideal and does not allow to fully benefit from the unique characteristics of such sensors. Spiking Neural Networks (SNN) models have been studied for several years as an interesting alternative to conventional Neural Networks both for their reduction of computational complexity in deep network topologies, and for their natural ability to support unsupervised and bio-inspired learning rules. In the context of interest of image processing, SNN are particularly suitable with event-based sensors and are therefore more suited to capture spatio-temporal regularities in a constant flow of events. The combination of event-based sensors and SNN networks makes it possible to considerably increase the energy efficiency of neural-based AI Dampfhoffer et al. (2023); Lemaire et al. (2023) while guaranteeing low reaction times in the context of real-time embedded processing such as autonomous systems, especially autonomous vehicles, drones, or satellites. Much work has been proposed in this framework but very little of it addresses the question of fusion of frame-based and event-based cameras.

2 Goals of the post-doctoral project

This research project is positioned in a new framework for image processing in a continuous flow of information from sensors integrated into embedded devices. Standard approaches to image processing and machine learning are based on the notion of frames derived from the way information is captured by traditional image sensors. However, the functioning of the biological retina and of the visual system more generally reminds us that this information is continuous and that other, more efficient, computation schemes are possible. The emergence of event-driven sensors in the 2000s has helped to challenge the sensors themselves, but without sufficiently modifying the algorithms that manipulate these nanoscale event flows. The reconstruction of intermediate frames continues to bring the neuromorphic community back into the well-established framework of conventional image processing.

In this project, we're seeking to free ourselves as much as possible from this representation, to take full advantage of the temporal dynamics of the scene in which the system (drone, vehicle, robot, satellite, etc.) is evolving. This paradigm shift requires us to master the theory of event-driven processing, to learn spiking neural networks that can effectively take advantage of their sparsity, and to adapt conventional image processing to this different nature of visual information, closer to that present in the brain.

Visual dynamics will be established through optical flow prediction and will be used to correct the visual perception of an embedded system. We will first focus on semantic segmentation tasks and seek to demonstrate that this computational paradigm can both increase processing throughput and reduce the overall energy consumption of embedded AI.

The project will draw on complementary research results from the **eBRAIN group from LEAT laboratory at Université Côte d'Azur** on SNN networks and neuromorphic architectures, and from the **Neuromorphic Group at the University of Groningen** on bio-inspired sensors for visual perception. The postdoc will be supervised both by Pr. Benoit Miramond from Université Côte d'Azur and Pr. Elisabetta Chicca from Groningen University.

3 State of the art

Alonso et al. Alonso and Murillo (2019) were pioneers in developing an Xception-based encoder-decoder model for semantic segmentation that used event data. Gehrig et al. Gehrig et al. (2020)

further enhanced this approach by integrating synthetic event data from video sources to refine the training process. Subsequent research expanded on these ideas, with Wang et al. Wang et al. (2021a) investigating event-to-image transformations, Sun et al. Sun et al. (2022) introducing unsupervised domain adaptation techniques, and Wang et al. Wang et al. (2021b) applying knowledge distillation methods. Hybrid models that combine spiking and artificial neural networks have also been developed, as demonstrated by Zhang et al. Zhang et al. (2023) and Das et al. Das Biswas et al. (2024). A significant advancement came from Patel et al. Patel et al. (2021), who designed a spiking neural network architecture focused on segmenting images with only two classes. Following this, Kim et al. Kim et al. (2022) created an encoder-decoder spiking neural network architecture with batch normalization through time, capable of handling a greater number of classes. Building on these foundations, Hareb et al. Hareb and Martinet (2024) introduced a fully spiking network, and Zhang et al. Zhang et al. (2024) recently proposed a spike-based hierarchical search method to optimize encoder architectures for both event streams and fused event-image inputs.

Recent advances in event-based optical flow estimation have transformed the field through increasingly sophisticated methodologies. The development began with Zhu et al. Zhu et al. (2018), who introduced EV-Flownet, a self-supervised model that segments events into nearby temporal bins and uses contrast maximization for motion compensation. Building on this, Ye et al. Ye et al. (2020) proposed an unsupervised approach to simultaneously estimate optical flow, depth, and egomotion. Moving beyond traditional event accumulation, Hagenaars et al. Hagenaars et al. (2021) and Cuadrado et al. Cuadrado et al. (2023) introduced SNNs to better capture the spatiotemporal structure of event data in continuous settings. In supervised approaches, Gehrig et al. Gehrig et al. (2021) adapted the RAFT model Teed and Deng (2020) for events, introducing correlation volumes to the event domain—a technique later refined by Hagenaars et al. Hagenaars et al. (2021), Paredes et al. Paredes-Vallés and De Croon (2021), and Shiba et al. Shiba et al. (2022). Further enhancing these methods, Paredes et al. Paredes-Vallés et al. (2023) and Liu et al. Liu et al. (2023) worked to improve accuracy and temporal resolution: Paredes used an iterative event-warping module to potentially capture nonlinear motion trajectories, while Liu added finer temporal correlation volumes. However, while these methods achieved high performance, the computation and storage requirements of correlation volumes, as seen in Gehrig et al. (2021) Liu et al. (2023), and later Gehrig et al. (2024)—which fit Bezier curves for nonlinear motion—scale poorly with input resolution, limiting use on memory-constrained devices. To address these constraints, Wu et al. Wu et al. (2024) introduced a sequential processing method that bypasses correlation volumes, achieving a low-latency, energy-efficient solution that remains effective even with large pixel displacements.

4 Organization of the postdoc position

The postdoc position will be organized as follows:

- formalization of the problem
- preparation of a bibliographic report
- first visit at Groningen to define the main features of the scientific approach
- develop the algorithm for event-based streaming perception with SNN
- experimentation on existing event-based datasets
- second visit at Groningen to collect data on bio-inspired sensors
- submission of publications
- duration: 18 months
- starting date: january 2026
- location: LEAT, Sophia-Antipolis

5 Skills

PhD degree in one of the following domains artificial intelligence, image processing, neuromorphic engineering, spiking neural networks.

Background and experience in machine-learning, spiking neural networks and/or embedded systems.

Strong motivation, team working, fluent in english spoken and written.

Programming skills in python, keras, pytorch or equivalent.

References

I. Alonso and A. C. Murillo. Ev-segnet: Semantic segmentation for event-based cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

J. Cuadrado, U. Rançon, B. R. Cottereau, F. Barranco, and T. Masquelier. Optical flow estimation from event-based cameras and spiking neural networks. *Frontiers in Neuroscience*, 17:1160034, 2023.

M. Dampfhoffer, T. Mesquida, A. Valentian, and L. Anghel. Are SNNs Really More Energy-Efficient Than ANNs? an In-Depth Hardware-Aware Study. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(3):731–741, June 2023. ISSN 2471-285X. doi: 10.1109/TETCI.2022.3214509. URL https://ieeexplore.ieee.org/document/9927729/?arnumber=9927729. Conference Name: IEEE Transactions on Emerging Topics in Computational Intelligence.

S. Das Biswas, A. Kosta, C. Liyanagedera, M. Apolinario, and K. Roy. Halsie: Hybrid approach to learning segmentation by simultaneously exploiting image and event modalities. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5964–5974, 2024.

D. Gehrig, M. Gehrig, J. Hidalgo-Carrió, and D. Scaramuzza. Video to events: Recycling video datasets for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3586–3595, 2020.

M. Gehrig, M. Millhäusler, D. Gehrig, and D. Scaramuzza. E-raft: Dense optical flow from event cameras. In *2021 International Conference on 3D Vision (3DV)*, pages 197–206. IEEE, 2021.

M. Gehrig, M. Muglikar, and D. Scaramuzza. Dense continuous-time optical flow from event cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

J. Hagenaars, F. Paredes-Vallés, and G. De Croon. Self-supervised learning of event-based optical flow with spiking neural networks. *Advances in Neural Information Processing Systems*, 34: 7167–7179, 2021.

D. Hareb and J. Martinet. Evsegsnn: Neuromorphic semantic segmentation for event data. *arXiv preprint arXiv:2406.14178*, 2024.

Y. Kim, J. Chough, and P. Panda. Beyond classification: Directly training spiking neural networks for semantic segmentation. *Neuromorphic Computing and Engineering*, 2(4):044015, 2022.

E. Lemaire, L. Cordone, A. Castagnetti, P.-E. Novac, J. Courtois, and B. Miramond. An Analytical Estimation of Spiking Neural Networks Energy Efficiency. In M. Tanveer, S. Agarwal, S. Ozawa, A. Ekbal, and A. Jatowt, editors, *Neural Information Processing*, pages 574–587, Cham, 2023. Springer International Publishing. ISBN 978-3-031-30105-6. doi: 10.1007/978-3-031-30105-6_48.

H. Liu, G. Chen, S. Qu, Y. Zhang, Z. Li, A. Knoll, and C. Jiang. Tma: Temporal motion aggregation for event-based optical flow. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9685–9694, 2023.

F. Paredes-Vallés and G. C. De Croon. Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3446–3455, 2021.

F. Paredes-Vallés, K. Y. Scheper, C. De Wagter, and G. C. De Croon. Taming contrast maximization for learning sequential, low-latency, event-based optical flow. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9695–9705, 2023.

K. Patel, E. Hunsberger, S. Batir, and C. Eliasmith. A spiking neural network for image segmentation. *arXiv preprint arXiv:2106.08921*, 2021.

S. Shiba, Y. Aoki, and G. Gallego. Secrets of event-based optical flow. In *European Conference on Computer Vision*, pages 628–645. Springer, 2022.

Z. Sun, N. Messikommer, D. Gehrig, and D. Scaramuzza. Ess: Learning event-based semantic segmentation from still images. In *European Conference on Computer Vision*, pages 341–357. Springer, 2022.

Z. Teed and J. Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 402–419. Springer, 2020.

L. Wang, Y. Chae, and K.-J. Yoon. Dual transfer learning for event-based end-task prediction via pluggable event to image translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2135–2145, 2021a.

L. Wang, Y. Chae, S.-H. Yoon, T.-K. Kim, and K.-J. Yoon. Evdistill: Asynchronous events to end-task learning via bidirectional reconstruction-guided cross-modal knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 608–619, 2021b.

Y. Wu, F. Paredes-Vallés, and G. C. De Croon. Lightweight event-based optical flow estimation via iterative deblurring. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14708–14715. IEEE, 2024.

C. Ye, A. Mitrokhin, C. Fermüller, J. A. Yorke, and Y. Aloimonos. Unsupervised learning of dense optical flow, depth and egomotion with event-based sensors. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5831–5838. IEEE, 2020.

H. Zhang, X. Fan, and Y. Zhang. Energy-efficient spiking segmenter for frame and event-based images. *Biomimetics*, 8(4):356, 2023.

R. Zhang, L. Leng, K. Che, H. Zhang, J. Cheng, Q. Guo, J. Liao, and R. Cheng. Accurate and efficient event-based semantic segmentation using adaptive spiking encoder–decoder network. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2024. doi: 10.1109/TNNLS.2024.3437415.

A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis. Ev-flownet: Self-supervised optical flow estimation for event-based cameras. *arXiv preprint arXiv:1802.06898*, 2018.