
Neuromorphic saliency detection of multiple objects of interest in an event based scene

Amélie Gruel^{*1} and Jean Martinet²

¹Laboratoire d'Informatique, Signaux, et Systèmes de Sophia Antipolis – Université Nice Sophia Antipolis (... - 2019), COMUE Université Côte d'Azur (2015 - 2019), Centre National de la Recherche Scientifique : UMR7271, Université Côte d'Azur, COMUE Université Côte d'Azur (2015 - 2019) – France

²Laboratoire d'Informatique, Signaux, et Systèmes de Sophia Antipolis – Université Nice Sophia Antipolis (1965 - 2019), Centre National de la Recherche Scientifique, Université Côte d'Azur – France

Résumé

The joint use of silicon retinas (Dynamic Vision Sensors, DVS) and Spiking Neural Networks (SNNs) is a promising combination for dynamic visual data processing. Both technologies have recently emerged separately about a decade ago from electronics and neuroscience communities, sharing many features: biological inspiration, temporal dimension, model sparsity, aim for a higher energy efficiency, etc.

However, traditional and neuromorphic computer vision models can have difficulties handling a great amount of data simultaneously while minimising their energy consumption, especially at a high temporal resolution. A recent study shows that in certain lightning conditions, high resolution event cameras produce data susceptible to temporal noise and with an increasingly high per pixel event rate, thus leading to the decreased performance of some traditional computer vision tasks (1). A remedy for such an issue could be found in event data downscaling (2) — however the trade-off between information retention and data reduction with existing methods is not yet ideal.

We thus believe that applying visual attention to selectively acquire relevant information is a more appropriate approach to optimise the on- and off-line processing of event data. In order to test this theory, we have implemented in a first iteration a neuromorphic spatio-temporal attention using adaptive mechanisms (3) on CPU, Loihi and SpiNNaker. This model detects regions with higher event density by using inherent SNN dynamics combined with online weight and threshold adaptation.

We are now presenting a new model allowing for the simultaneous detection of multiple regions of interest in event data, solely relying on the intrinsic dynamics of SNN. This novel approach is promising for many relevant computer vision tasks such as simultaneous object classification. This model can detect a limited number of regions of interest at a time, and we will work towards a model that intrinsically adapts to the number of objects of interest in the scene.

(1) Gehrig, D. and Scaramuzza, D. (2022). Are high-resolution cameras really needed? *arXiv*.

*Intervenant

- (2) Gruel A. et al (2022). Event data downscaling for embedded computer vision. *VIS-APP*.
- (3) Gruel, A. et al (2022). Neuromorphic event-based spatio-temporal attention using adaptive mechanisms. *AICAS*.

Mots-Clés: spiking neural networks, event based camera, visual attention, neuromorphic