Neuromorphic Stereo Vision with Event Cameras
PhD proposal

- Title: Neuromorphic Stereo Vision with Event Cameras
- Supervisor: Prof. Jean Martinet, i3S/Polytech/UCA, jean.martinet@univ-cotedazur.fr
- Potential co-supervisor: /
- Laboratory and research group: I3S-SPARKS

Summary

Event cameras are bio-inspired vision sensors that work in radically different ways from traditional cameras. Instead of capturing images at a fixed rate, they measure changes in brightness per pixel asynchronously. It results in a flow of events, which encode the instant, location and sign of brightness changes. Event sensors show exceptional properties compared to traditional cameras: very high dynamic range (140 dB against 60 dB), high temporal resolution (order of microseconds), low latency, low energy consumption and no motion blur (see this example). Therefore, these sensors bring a great potential for computer vision in challenging scenarios. Beyond this great potential, stereopsis enables depth perception of the world, which is a key feature for both artificial and human visual processing systems. Besides, depth is an essential requirement for many practical applications, ranging from fine object manipulation in robotics, to autonomous driving for vehicles.

Inspired from biology, Spiking Neural Networks (SNN) are a special class of artificial neural networks in that they can work continuously and function more like the brain. SNN does not rely on stochastic gradient descent and backpropagation. Instead, neurons are connected through synapses, that implement a learning mechanism inspired from biology: it rests upon the "Spike-Timing-Dependent Plasticity" rule (STDP). SNN show many interesting features for processing event data, such as their asynchronicity, and their implementation on ultra-low-power neuromorphic hardware.

In this PhD proposal, we wish to design and implement an innovative neuromorphic model for stereo matching using event cameras. The supervisor belongs i3S lab, in the topics of bio-inspired machine learning and computer vision. This PhD proposal relates to the EU program APROVIS3D, that started in April 2020, and that targets embedded bio-inspired machine learning and computer vision, with an application to autonomous drone navigation. The work will be carried out in close collaboration with a PhD student who started on October 2020.

Keywords: Neuromorphic computing, Spiking neural nets, Event camera, Stereo matching, Computer Vision.

Description

Convolutional Neural Networks (CNN) are a brain inspired technique in artificial intelligence. They find applications in numerous areas, including self-driving cars, data analysis, commercial recommender systems and many more. In less than a decade, deep CNN such as Inception and VGG-16 have successfully pulled state-of-the-art classification performances to new levels, especially on challenging computer vision benchmarks like ImageNet. The availability of both tremendous amounts of annotated data and huge computational resources have enabled remarkable progress. Therefore, this success comes with substantial human cost required for manually labeling data, and energy cost required for inference, despite most recent advances in parallel digital architectures. Namely, training deep CNN requires tremendous amounts of power. For instance, ResNet has been trained for 3 weeks on an 8-GPU server, which is equivalent to a power consumption of about 1 GWh. More generally, worldwide data centers in general require a power of about 1 PW, which is equivalent to 4% of GHG emissions, which is over air transportation. Forecasts plan that this figure will double every 4 years.

On the other hand, the human brain has the ability to perform cognitive tasks with unrivalled computational and energy efficiency. It is believed that one major factor of this efficiency is the fact that information is represented by action potentials (or spikes) at analog –not discrete– times, in a sparse way. Inspired from biology, Spiking Neural Networks (SNN) are a special class of artificial neural networks in that they can work continuously and function more like the brain [Maass, 1997] [Ponulak, 2011] [Paugam-Moisy, 2012]. Spiking neurons communicate by sequences of spikes. Contrary to formal neurons, spiking neurons do not fire at each propagation cycle, but rather fire only when their activation level (or membrane potential, an intrinsic quality of the neuron related to its membrane electrical charge) reaches a specific threshold value. When a neuron fires, it generates a non-binary signal that travels to other neurons, which in turn increases their potentials. The activation level either increases with incoming spikes, or decays over time. SNNs have been shown to be computationally more efficient than standard rate-coding networks. In particular, they are more energy efficient if implemented on neuromorphic hardware. Neuromorphic hardware implementing SNN can be built with CMOS technology, and typically uses low power (under the threshold voltage), enabling to reduce energy dissipation by several orders of magnitude, compared to standard digital architectures [Merolla, 2014] [Desbief, 2015]. Regarding inference, SNN does not rely on stochastic gradient descent and backpropagation. Instead, neurons are connected through synapses, that implement a learning mechanism inspired from biology: it rests upon the “Spike-Timing-Dependent Plasticity”, a rule that updates synaptic weights (strength of connections) according to causal links observed between presynaptic and postsynaptic spikes. This updating rule reinforces incoming connections that cause the neuron to fire. Therefore, the learning process is intrinsically not supervised, and can be successfully used detect patterns in data in an unsupervised manner [Bichler, 2012] [Beyeler, 2013]. SNN are little used, yet there is an increasing interest in using such type of neural network in machine learning. A number of open issues and questions need to be addressed, such as the design of an efficient SNN topology (convolutional? layered? recurrent?), the understanding and control of the learning process with parameter tuning, the right way to input supervision (during or after inference?), the input coding (convert input data values to spike trains) and output decoding (interpret the output spikes).

SNN show many interesting features for processing event data, such as their asynchronicity, and their implementation on ultra-low-power neuromorphic hardware [Verzi, 2018] [Pei, 2019] [Roy, 2019] [Taherkhani, 2020]. Moreover, such implementations are needed to go beyond the limitations of Von Neumann architectures and to tackle the end of Moore law. And yet, a number of challenges lie ahead before they become a realistic alternative to deep CNN. Recent work shows that SNN are competitive with the state-of-the-art in computer vision on “easy” datasets such as MNIST (handwritten digits) [Diehl, 2015] [Kheradpisheh, 2018] [Falez, 2019].

Beside static images, because of their asynchronous operation principle, SNN are allegedly likely to handle well temporal data such as video. Event-based cameras (or silicon retinas) bring a new vision paradigm by mimicking the biological retina. Instead of measuring the intensity of every pixel in a fixed time interval, it reports events of significant pixel intensity changes. Every such event is represented by its position, sign of change, and timestamp, accurate to the microsecond. Because of their asynchronous operation principle, they are a natural match for SNN [Steffen, 2019] [Paredes, 2019] [Oudjail, 2019]. State-of-the-art approaches in machine learning provide excellent results for vision tasks with standard cameras, however, asynchronous event sequences require special handling, and spiking networks can take advantage of this asynchrony.

Moreover, conventional video sensors record the entire image with a given rate and resolution. The original rationale for sensing a scene this way is that the transmission or recording is intended to be viewed by a human observer who may be looking closely at any part of the moving image. Frame-based video contains a huge amount of redundant data and requires enormous computational power to process. As stated in [Hopkins, 2018], biological vision sensors, however, are quite different from frame-based cameras. They do not sample images at a uniform rate, nor at a uniform resolution. The human eye has a small high-resolution region (the fovea) in the center of the field of vision, and a much larger peripheral vision, which has much lower resolution, combined with an increased sensitivity to movement. Therefore, limited resources are deployed to extract the most salient information from the scene without wasting energy capturing the entire scene at the highest resolution. Furthermore, the human eye is primarily sensitive to changes in the luminance falling on its individual sensors. These changes are processed by layers of neurons in the retina through to the retinal ganglion cells that generate action potentials, or spikes, whenever a significant change is detected. Then these spikes propagate through the optic nerve to the brain. This approach focuses the resources on the areas of the image that convey most useful information such as edges and other details. Given the core objective of computer vision systems, it seems natural to sense the world with bio-inspired sensors. Moreover, primates and other mammals are given the ability to compute depth information from views acquired simultaneously from different points in space with stereopsis, which is a fundamental feature in environment 3D sensing.

Bio-inspired models from binocular vision have also been used to solve the event-based stereo correspondence problem such as in [Oswald, 2017] [Dikov, 2027] [Tulyakov, 2019] [Rissi, 2020]. These approaches often use several populations of neurons, such as retinal cell attached to event cameras outputs, and coincidence detectors for each polarity (direction of luminance variations), and disparity detectors that polls responses from the coincidence detector neurons using both excitatory and inhibitory connections. In this PhD proposal, after a study of existing neuromorphic methods (namely by Osswald, Dikov, Tulyakov, and Risi), we will start by selecting and implementing one method to obtain a functional prototype to be tested with existing stereo event datasets such as MVSEC, as well as a pair of Prophesee cameras (that will be provided) for live input. Then we will extend the prototype using advanced features and stereo matching techniques, e.g., HOTS [Lagorce, 2017] and PatchMatch Stereo [Bleyer, 2011].

Candidates must hold a Master degree in Computer Science, Computer Engineering, Applied Mathematics or a related field. Experience in image processing, computer vision, or machine learning is a plus. In addition, candidates should have the following skills: good proficiency in spoken and written English, scientific writing, and programming in Python/C++.

References:

[Beyeler, 2013] Michael Beyeler, Nikil D. Dutt et Jeffrey L. Krichmar : Categorization and decision-making in a neurobiologically plausible spiking network using a STDP-like learning rule. Neural Networks 48 (2013) 109– 124.
[Bleyer, 2011] Bleyer, Michael, Christoph Rhemann and Carsten Rother. “PatchMatch Stereo - Stereo Matching with Slanted Support Windows.” BMVC (2011). [Bichler, 2012] Olivier Bichler, Damien Querlioz, Simon J. Thorpe, Jean-Philippe Bourgoin, Christian Gamrat : Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. Neural Networks 32 (2012) 339–348.
[Desbief, 2015] Desbief, Simon ; Kyndiah, Adrica ; Guérin, David ; Gentili, Denis ; Murgia, Mauro ; Lenfant, Stéphane ; Alibart, Fabien ; Cramer, Tobias ; Biscarini, Fabio ; Vuillaume, Dominique/ Low voltage and time constant organic synapse-transistor. Organic Electronics, June 2015, Vol.21, pp.47-53.
[Dikov et al., 2017] G. Dikov, M. Firouzi, F. Röhrbein, J. Conradt, and C. Richter, ‘Spiking Cooperative Stereo-Matching at 2 ms Latency with Neuromorphic Hardware’, 2017, doi: 10.1007/978-3-319-63537-8_11.
[Falez, 2019] Unsupervised Visual Feature Learning with STDP: How Far are we from Traditional Feature Learning Approaches? Pattern Recognition, 2019.
[Kheradpisheh, 2018] STDP-based spiking deep convolutional neural networks for object recognition. Neural Networks (99:56–67), 2018.
[Lagorce, 2017] X. Lagorce, G. Orchard, F. Gallupi, B. E. Shi, and R. Benosman, “HOTS: A hierarchy of event-based time-surfaces for pattern recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 7, pp. 1346–1359, 2017.

[Maass, 1997] Networks of spiking neurons: the third generation of neural network models. W Maass. Neural Networks 10 (9), 1997
[Merolla, 2014] A million spiking-neuron integrated circuit with a scalable communication network and interface. P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha. Science, vol. 345, pp. 668–673, Aug. 2014.
[Osswald, 2017] M. Osswald, S.-H. Ieng, R. Benosman, and G. Indiveri, 'A spiking neural network model of 3D perception for event-based neuromorphic stereo vision systems', Scientific reports, 2017, doi: 10.1038/srep40703.
[Paugam-Moisy, 2012] Hélène Paugam-Moisy, Sander M. Bohte. Computing with Spiking Neuron Networks. G. Rozenberg, T. Back, J. Kok. Handbook of Natural Computing, Springer-Verlag, pp.335-376, 2012,
[Pei, 2019] Towards artificial general intelligence with hybrid Tianjic chip architecture. Jing Pei, Lei Deng, Sen Song, Mingguo Zhao, Youhui Zhang, Shuang Wu et al. Nature, 572, 2019
[Ponulak, 2011] Filip Ponulak, Andrzej Kasiński. Introduction to spiking neural networks: Information processing, learning and applications. Acta Neurobiol Exp (Wars). 2011;71(4):409-33.
[Risi et al., 2020] N. Risi, A. Aimar, E. Donati, S. Solinas, and G. Indiveri, 'A Spike-Based Neuromorphic Architecture of Stereo Vision', Front. Neurorobot., vol. 14, 2020, doi: 10.3389/fnbot.2020.568283.
[Roy, 2019] Towards spike-based machine intelligence with neuromorphic computing. Kaushik Roy, Akhilesh Jaiswal & Priyadarshini Panda, Nature, 575, Nov. 2019.
[Steffen, 2019] Steffen, L., Reichard, D., Weinland, J., Kaiser, J., Roennau, A., Dillmann, R., Neuromorphic Stereo Vision: A Survey of Bio-Inspired Sensors and Algorithms, Front. Neurorobot. (2019) 13:28.
[Taherkhani, 2020] A review of learning in biologically plausible spiking neural networks. Aboozar Taherkhani, Ammar Belatreche, Yuhua Li, Georgina Cosma, Liam P. Maguire, T.M. McGinnity. Neural Networks, 122, Feb. 2020.
[Tulyakov, 2019] Tulyakov, S., Fleuret, F., Kiefel, M., Gehler, P., Hirsch., M., Learning an event sequence embedding for dense event-based deep stereo, IEEE Int. Conf. Computer Vision (ICCV), 2019.
[Verzi, 2018] Computing with Spikes: The Advantage of Fine-grained Timing. Stephen J. Verzi, Fredrick Rothganger, Ojas D. Parekh, Tu-Thach Quach, Neural Computation, 30(10), October 2018.
[Prophessee] URL: https://www.prophesee.ai
[MVSEC] URL : https://daniilidis-group.github.io/mvsec/