Title: Neuromorphic Visual Odometry for Intelligent Vehicles with a Bio-inspired Vision Sensor

Position: Thesis
Supervisor: Jean Martinet
Co-supervisor: Andrew Comport
Laboratory: I3S (Teams SPARKS and SIS)
Link to detailed description: http://i3s.unice.fr/jmartinet/en/node/14

Summary

This thesis aims at exploiting biologically devised 'short cuts' used by insects with small brains and relatively simple nervous systems to see and perceive their world in real-time. The objective is to develop a biologically-inspired omni-directional event camera model to perform *real-time* ego-motion estimation and environment mapping. Spiking neural networks (SNN) approaches, that are particularly adapted to biologically-inspired event-cameras, will be developed to exploit asynchronous events in real-time for autonomous navigation. A novel panoramic stereo event camera system will be developed and arranged in a spherical configuration inspired by insects. Algorithms will be devised to exploit the 360-degree field of view, high-frequency event streaming, in order to perform visual odometry using landmarks. Bio-inspired models from binocular vision have also been used to solve the event-based stereo correspondence problem and spiking neural networks are natural match for event-based cameras due to their asynchronous operation principle. The main applicative goal of this work will be to exploit the high-temporal resolution, high-dynamic range and the low-power consumption of event cameras to perform high-speed robotics applications such as drone navigation. The major challenge will be to exploit the full high-speed potential of event cameras by redefining classic real-time spherical RGB-D SLAM approaches within an asynchronous framework. This can be divided into three goals: develop mathematical models for asynchronous multi-event cameras; develop SNN approaches for spatio-temporal stereo from events; develop high-speed visual odometry techniques for real-time navigation and mapping. The long-term goal of this proposed project is to study new paradigms and concepts for *real-time* spatial intelligence in the conditions of extreme lighting, exploiting the new type of visual sensor.

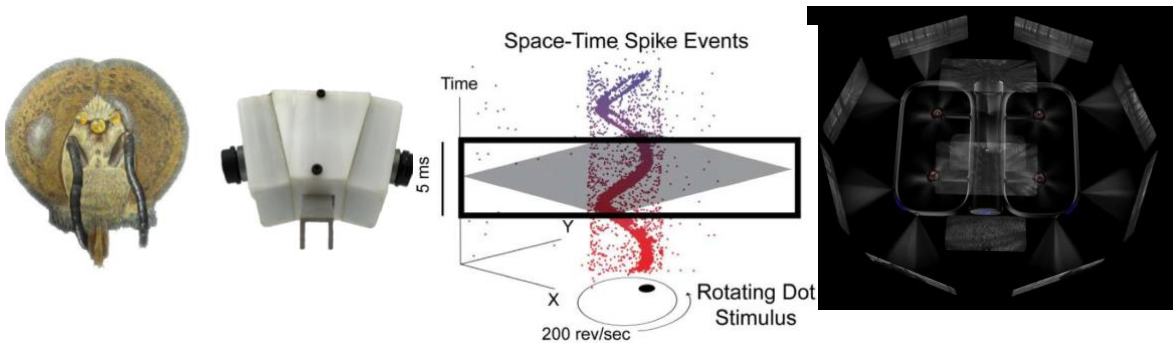
This project has been granted initial funding in 2020 by the UNS/UCA CSI (BISSAI project) and I3S internal call for project. The grant has already funded a pair of stereo event cameras and a Master studentship.

Detailed Description

Research in understanding the elegant methods used by animals with small brains and relatively simple nervous systems to see and perceive their world, and to navigate in it have revealed how flying insects are extremely negotiate narrow gaps, regulate the height and speed of flight, estimate distance flown, and orchestrate smooth landings [21, 22]. 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 time-stamp, accurate to the microsecond. Because of their asynchronous operation principle, they are a natural match for SNN [18, 23]. 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.

Autonomous cars and mobile drones, on the other hand, are currently in phase with becoming mainstream commercial platforms offering a wide range of advantages. An important theoretical aspect of these systems is the *real-time* control and navigation of these autonomous robots and a predominant sensor in achieving this goal is the visual sensor. These real-time applications have a fundamental need for the high temporal resolution and low latency ego-motion estimation which has, until now, been fulfilled by combined vision and inertial sensors. The former being low frame-rate and the latter being subject to drift. Event cameras not only provide an alternative solution to high-speed visual-inertial odometry by they have the added advantages of low power consumption, low motion blur and high dynamic range sensing makes this sensor modality unavoidable for the future of intelligent and agile autonomous systems. Another fundamental requirement is for safe and robust autonomous systems. The majority of commercial systems such as autonomous cars (Renault, Google, Uber, …) or drones (Skydio) therefore target 360-degree aware systems composed of multiple camera and radar systems providing both color and depth information (RGB-D) – see Figure 1.

This thesis therefore aims at develop *real-time* ego-motion estimation and environment mapping approaches based on state-of-the-art spiking CNN techniques for a new panoramic spherical RGB-D event camera system that is capable of performing 360-degree color and depth sensing. The major challenge will be to exploit the full high-speed potential of event cameras by redefining classic real-time spherical RGB-D SLAM approaches within an asynchronous framework. The long-term goal of this proposed project is to study new paradigms and concepts for *real-tim*e spatial intelligence in the conditions of extreme lighting, exploiting the new type of visual sensor.



(a) (b) (c)

Figure 1. Spherical sensing. (a) The compound eyes of a honey bee (M.J. Roberts) (b) Space-time spike events (c) Skydio spherical RGB-D drone sensors.

Related works

The main goal of this work will be to exploit the high-temporal resolution of event cameras and their low power consumption [20] to perform high-speed robotics applications. Whilst this is the benefit found in most state-of-the-art event camera approaches, the advantage should be even more apparent in multi-camera systems that provide a much higher volume of data compared to a single camera.

Few works exist on 360-degree event-camera stereo, and those that do perform panoramic imaging by rotating the cameras [4, 5]. This particular approach has the intrinsic disadvantage of trading off

the high temporal resolution of event cameras for panoramic vision since the cameras must wait to perform one full rotation before acquiring a complete 360-degree scan.

Significantly more research has been carried out on the problem of stereo event cameras [7,8,9]. Stereo triangulation allows to perform depth estimation and 3D reconstruction using two or more cameras with a rigid baseline and synchronized using a common clock. Usually points are first matched between images and geometric triangulation is then performed. Event cameras are similar to rolling shutter cameras [3] in that movement may occur asynchronously across the image. Stereo event camera approaches differ mainly in the matching approach, some reconstruct first a classic image frame by accumulating events over time [10, 11] and then use classic stereo matching approaches. Others exploit the temporal correlations between events across sensors [12, 13].

Bio-inspired models from binocular vision have also been used to solve the event-based stereo correspondence problem such as in [16, 17]. Spiking Neural Networks are a special class of artificial neural networks, where neurons communicate by sequences of asynchronous spikes. Therefore, they are a natural match for event-based cameras due to their asynchronous operation principle [18]. State-of-the-art approaches in machine learning provide promising results for 3D mapping, however, asynchronous event sequences require special handling, and spiking networks can take advantage of this asynchrony. Besides, asynchronous convolutions [19] must be considered if one is to not lose sight of the goal to perform high speed efficient processing for navigation and control.

The I3S robot vision group is specialized in real-time visual localization and mapping (visual SLAM [2]) and has already a solid background on spherical RGB-D sensing [1], High Dynamic Range mapping of large-scale scenes [3] and high-speed rolling shutter ego-motion estimation [4]. They have also worked on a generalized camera model for visual servoing which is of interest to the proposed project [5].

Expected results

The aim of this exploratory project is to develop a multi-event-camera system for visual localization, mapping and semantic classification [1, 2, 3, 4].

The first phase of this PhD will involve developing a mathematical model for a multi-event camera device for capturing raw 360 degree event data that is parametrized according to varying configurations including: the physical layout of the sensors vertical or in ring format; the baseline or interocular distance; the camera overlap; divergence or convergence; the uncertainty of 3D reconstruction; calibration and particularly the synchronization of the sensors. The generalized camera model [15] will be extended to asynchronous event cameras by augmenting each 3D viewing ray with a time-stamp and event information.

In a second step, "machine learning" approaches will be studied to perform 3D dense matching and reconstruction by exploiting spatio-temporal convolutions. This will first involve implementing and comparing with state-of-the-art matching and reconstruction approaches developed for event cameras.

The next step will involve performing “high-speed” localization and mapping by exploiting the asynchronous data real-time. Localization will be considered in both a classic non-linear estimation context along with recent convolutional network approaches. The main goal will be to exploit the high-temporal resolution of the sensor, along with the asynchronous nature of the events.

In a final part, high-level machine learning approaches will be adapted to enrich the mapping and reconstruction by simultaneously performing semantic classification of information relevant to the ego-motion of the sensor.

PhD supervisors and Work environment

The principal supervisor from the I3S, Professor Jean Martinet, is newly appointed at the I3S and is member of the SPARKS team. His research is focused on the study of spiking neural networks and their relation to event cameras for vision. He was formerly with the University of Lille (France) during 12 years, where he was the head of a research group in Computer Vision at CRIStAL, the largest IT French lab in the north of Paris.

The co-supervisor from the I3S, Andrew Comport, is a visual SLAM specialist. He holds a permanent appointment as a researcher with the CNRS. His research is focused on the fields of robot learning and machine vision and in particular on real-time spatial AI involving semantic and 3D mapping, deep learning, tracking, localization, real-time computation and visual servoing (Real-time Spatial AI). The international recognition of this work is visible with over 50 international publications and more than 2080 citations. This recognition is also attested by an overall Best Paper at the flagship international IEEE/RSJ conference on Robotics (IROS 2013). In 2005 he co-founded the start-up PIXMAP which was successfully acquired by one of the 'Big 5' from the Silicon Valley. This success stems from an active history industrial transfer along with 2 patents and 6 software patents (APP). He is currently Associate Editor IEEE RA-L and the IEEE International Conference on Robotics and Automation (ICRA).

The I3S laboratory is the largest information and communication science laboratory in the French Riviera. It is composed of nearly 300 people with approximately 100 professors or associate professors, 20 CNRS researchers and 13 INRIA researchers, along with 20 technical and administrative staff. The I3S robot-vision group also has a collaboration with Robert Mahony from the ANU and Tom Drummond from Monash university in Australia until 2024 via a CNRS international laboratory (LIA) which could provide the basis for collaboration on this topic. Jean Martinet is responsible for the European CHIST-ERA project (starting during spring 2020) on the theme of bio-inspired machine learning for stereo event-based vision, using mixed analog-digital hardware. This project provides a complementary research axis to the current proposal and collaboration will provide a solid basis for furthering research on event cameras.

[1] A Spherical Robot-Centered Representation for Urban Navigation, Maxime Meilland. Andrew I. Comport, Patrick Rives. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Oct 2010, Taipei, Taiwan.

[2] On unifying key-frame and voxel-based dense visual SLAM at large scales, Maxime Meilland, Andrew I. Comport, International Conference on Intelligent Robots and Systems, 2013, Tokyo, Japan.

[3] A unified rolling shutter and motion blur model for 3d visual registration, Maxime Meilland, Tom Drummond, Andrew I. Comport, Proceedings of the IEEE International Conference on Computer Vision, 2013, Sydney, Australia. pp.2016–2023

[4] Active High Dynamic Range Mapping for Dense Visual SLAM, Christian Barat, Andrew I. Comport, IEEE/RSJ International Conference on Intelligent Robots and Systems, Sep 2017, Vancouver, Canada.

[5] A Visual Servoing Model for Generalised Cameras: Case study of non-overlapping cameras A.I. Comport, R. Mahony, F. Spindler. IEEE Int. Conf. on Robotics and Automation, ICRA'11, 2011, Shanghai, China, China. pp.5683-5688

[6] Belbachir, A., Pflugfelder, R., Gmeiner, P., A Neuromorphic Smart Camera for Real-time 360deg distortion-free Panoramas, IEEE Conference on Distributed Smart Cameras, 2010.

[7] S. Schraml, A. N. Belbachir, Bischof, H., An Event-Driven Stereo System for Real-Time 3-D 360° Panoramic Vision, IEEE Trans. Ind. Electron., 63(1):418-428, 2016.

[8] 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, Jul. 2017.

[9] Zhou, Y., Gallego, G., Rebecq, H., Kneip, L., Li, H., Scaramuzza, D., Semi-Dense 3D Reconstruction with a Stereo Event Camera, European Conf. Computer Vision (ECCV), 2018.

[10] Schraml, S., Belbachir, A. N., Milosevic, N., Schon, P., Dynamic stereo vision system for real-time tracking, IEEE Int. Symp. Circuits and Systems (ISCAS), 2010, pp. 1409-1412.

[11] Kogler, J., Sulzbachner, C., Humenberger, M., & Eibensteiner, F. Address-event based stereo vision with bio-inspired silicon retina imagers. In Advances in Theory and Applications of Stereo Vision, pp. 165–188, 2011.

[12] Kogler, J., Humenberger, M., & Sulzbachner, C. Event-based stereo matching approaches for frameless address event stereo data. In International Symposium on Advances in Visual Computing (ISVC), pp. 674–685, 2011.

[13] Camunas-Mesa, L. A., Serrano-Gotarredona, T., Ieng, S. H., Benosman, R. B., & Linares-Barranco, B. On the use of orientation filters for 3D reconstruction in event-driven stereo vision. Frontiers in Neuroscience, 8, 48, 2014.

[14] Maqueda et. al. CVPR 2018. Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars.

[15] Macanovic, M., Chersi, F., Rutard, F., Ieng, S.-H., Benosman, R., When Conventional machine learning meets neuromorphic engineering: Deep Temporal Networks (DTNets) a machine learning framework allowing to operate on Events and Frames and implantable on Tensor Flow Like Hardware, arXiv: 1811.07672, 2018.

[16] 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, vol. 7, no. 1, Jan. 2017.

[17] 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.

[18] 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.

[19] Scheerlinck, C., Barnes, N., Mahony, R., Asynchronous Spatial Image Convolutions for Event Cameras, IEEE Robotics and Automation Letters (RA-L), 4(2):816-822, Apr. 2019.

[20] FINATEU, Thomas, BARRANCO, Bernabé LINARES, GOTARREDONA, Teresa SERRANO, *et al. Pixel circuit for detecting time-dependent visual data*. U.S. Patent Application No 16/343,720, 5 sept. 2019.

[21] Neural basis of forward flight control and landing in honeybees. Ibbotson MR, Hung YS, Meffin H, Boeddeker N, Srinivasan MV. Sci Rep. 2017 Nov 6;7(1):14591.

[22] Going with the flow: a brief history of the study of the honeybee's navigational 'odometer'. Srinivasan MV. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 2014 Jun;200(6):563-73. doi: 10.1007/s00359-014-0902-6. Epub 2014 Apr 17.

[23] Veïs Oudjail, Jean Martinet: Bio-inspired Event-based Motion Analysis with Spiking Neural Networks. VISIGRAPP (4: VISAPP) 2019: 389-394