

Workshop on Emerging Visual sensors

Detailed Program (abstracts included)

Venue : Delegation CNRS Ivry-sur-Seine
Metro 7: Porte de Choisy
Address: Delegation CNRS 27 rue P. Bert Ivry-sur-Seine
<https://goo.gl/maps/TBSfB1Yh72E2>

Authors are encouraged to make their presentations in English.

Registration is compulsory :
<http://www.gdr-isis.fr/index.php?page=reunion&idreunion=385>

- 09:30 Welcome by F. Ruffier, G. Caron, P. Vasseur
- 09:40 Sio-Hoi Ieng (Institut de la vision, Paris) Invited talk
- 10:20 Daniel Matolin (Prophesee) Regular talk
- 10:45 Coffee break (20 minutes)
- 11:05 Guillaume Debat, Michel Paindavoine (Cerco Toulouse / GST Dijon) Regular talk
- 11:30 Tushar Chauhan et al. (Cerco, Toulouse) Short talk with additional poster
- 11:45 Veis Oudjail et al. (Cristal, Lille Univ.) Short talk with additional poster
- 12:00 Lunch
- 13:20 Interactive poster session (30 minutes)
- 13:50 Tobi Delbruck (INI Zurich, CH) Invited talk
- 14:30 Stéphane Chevobbe (CEA-List Nano-Innov) Regular talk
- 14:55 Camille Simon-Chane et al. (ETIS Cergy/ Institut de la vision Paris) Regular talk
- 15:20 Coffee break + Interactive poster session (30 minutes)
- 15:50 Guido de Croon (TU Delft, NL) Invited talk
- 16:30 Kévin Hoang et al. (Thales GTS-SIX / ETIS Cergy) Regular talk
- 16:55 Closing

Posters:

- Victor Boutin et al. (INT/ISM, Aix-Marseille Univ)
- Stéphane Mancini (TIMA, Grenoble INP/Univ.)
- Tushar Chauhan et al. (Cerco, Toulouse)
- Veis Oudjail et al. (Cristal, Lille Univ.)

INVITED TALKS

9:40 Sio-Hoi IENG*Vision Institute, Sorbonne Universités - CNRS - INSERM, Paris, FR***Title: Event-based sensing: “One event, one update”***Abstract:*

The event-based vision sensors are recently attracting machine vision community, however the foundations has been built through almost a decade long research by the vision and natural computation groupe when the DVS from INI was made to academics. In this presentation, an panorama to the event-based vision sensing from the signal sampling to its processing through a set of examples is given. The objective of this presentation is also to show an alternative way of tackling visual signal processing: we have paved new paths rather than following the comforting scheme of converting events into a form (i.e. frames) that is complying with “modern” computer vision techniques. We are crafting from scratch methodologies that are natively adapted to the event-based sensing in order to build an entire new paradigm of its own.

Biography: Associate Pr. affiliated to the vision institute since 2010. His initial research domain was omnidirectional machine vision and have changed to event-driven vision sensing since 2008. The covered research spans from low level event-based signal processing to higher level machine vision tasks such as colored vision segmentation, features extraction, pose estimation, pattern recognitions and some machine learning.

13:50 Tobi DELBRUCK*Institute of Neuroinformatics, University of Zurich and ETH Zurich, CH***Title: History and recent developments of dynamic vision sensor event cameras**

Abstract: This talk is about neuromorphic “silicon retina” event cameras that output asynchronous spike-events modeling a simplified transient pathway of biological retinas, sometimes including conventional synchronous or asynchronous grey frames. These dynamic vision sensors (DVS) offer advantages for real-world vision problems in terms of latency, dynamic range, temporal resolution, and post-processing cost. There are opportunities for theoretical and practical developments of algorithms and hardware architecture for dynamic vision applications. The talk will include several live demonstrations of a recent sensor, including high speed stabilization and visual classification adaptively driven by DVS.

<http://sensors.ini.uzh.ch>

<https://www.youtube.com/user/tobidelbruck>

Biography: Tobi Delbruck (IEEE M’99–SM’06–F’13) received the B.Sc. degree in physics from University of California in 1986 and PhD degree from California Institute of Technology in 1993. Currently, he is a Professor of Physics and Electrical Engineering with the Institute of Neuroinformatics, University of Zurich and ETH Zurich, where he has been since 1998. The Sensors group focuses on neuromorphic sensory processing and efficient hardware deep neural networks.

15:50 **Guido DE CROON**

TU Delft, Delft, NL

*Title: **Event-based vision for small drones***

Abstract: Tiny, light-weight drones are very safe for flight around humans, navigating in indoor, narrow spaces, while still capturing the information necessary for the application such as warehouse stock tracking or search-and-rescue. For such applications, the tiny drones will have to fly completely by themselves. This is very challenging, since they are very restricted in terms of sensing, processing power, and memory. For the solution, bio-inspiration is key. In the talk, I will present the research we performed at TU Delft to achieve fully autonomous flight of tiny, light-weight swarms of drones. In particular, I will delve into the relevance of event-based cameras and spiking neural networks for these drones, and present some of our work on this topic.

Biography: Guido De Croon received his M.Sc. and Ph.D. in the field of Artificial Intelligence (AI) at Maastricht University, the Netherlands. His research interest lies with computationally efficient algorithms for robot autonomy, with an emphasis on bio-inspired methods of computer vision. Since 2008 he has worked on algorithms for achieving autonomous flight with small and light-weight flying robots, such as the DelFly flapping wing MAV. In 2011-2012, he was a research fellow in the Advanced Concepts Team of the European Space Agency, where he studied topics such as optical flow based control algorithms for extraterrestrial landing scenarios. Currently, he is associate professor at TU Delft and scientific lead of the Micro Air Vehicle lab (MAV-lab) of Delft University of Technology.

REGULAR TALKS

11:05 **Guillaume Debat**¹, **Robin Baurès**¹, **Benoit Cottureau**¹, **Timothée Masquelier**¹, **Tushar Chauhan**¹, **Clément Joly**², **Olivier Brousse**², **Michel Paindavoine**²

¹ *CerCo, Université de Toulouse, CNRS, UPS, Toulouse FR*

² *GlobalSensing Technologies, Dijon, FR*

Title: Analyse de trajectoire 3D à base de caméras à spike

Abstract: Dans le cadre du projet ANR “Beating Roger Federer”, le laboratoire CERCO a pour but de reproduire les performances perceptives et d’anticipation que développent les experts à travers leur apprentissage. Nous souhaitons développer un modèle d’apprentissage visuel bio-inspiré afin de le comparer à la perception humaine. Celui-ci devra être capable de déterminer si une balle est “in” ou “out” et dans un deuxième temps, de prédire le lieu et la date d’arrivée de la balle en fonction d’une portion de sa trajectoire. L’acquisition des images et la génération des spikes est réalisée par une caméra intelligente conçue par GlobalSensing Technologies. Cette caméra est basée sur un imageur CMOS Aptina et d’un circuit Xilinx Zynq 7020. Dans le cas du projet, la caméra, configurée dans un format 128x120 pixels avec un débit de 240fps, génère en temps réel des spikes en considérant trois échelles d’images pré-filtrées par le FPGA au sein du circuit Zynq. Ces spikes sont ensuite traités par un réseau de neurones impulsionnels bio-inspirée qui simule le traitement du cortex visuel, avec une règle d’apprentissage non-supervisé de Spike-Timing-Dependent Plasticity (STDP, e.g., Masquelier & Thorpe, 2007 ; Masquelier 2012). Les performances de ce réseau sont ensuite évaluées. Les premiers résultats obtenus permettent d’établir un réseau de neurone sélectifs à des directions de déplacement simples.

References:

Masquelier, T., & Thorpe, S.J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3, e31.

Masquelier, T. (2012). Relative spike time coding and STDP-based orientation selectivity in the early visual system in natural continuous and saccadic vision: a computational model. *Journal of Computational Neuroscience*, 32, 425-41.

14:30 **Stéphane CHEVOBBE**

CEA-List Nano-Innov, Palaiseau, FR

Title: A 3D Stacked Smart Vision Chip for Low Latency Image Processing

Titre: Un capteur d’image intelligent en technologie d’intégration 3D faible latence dédié au traitement d’image

Abstract: Nous avons fabriqué un imageur intelligent, appelé RETINE en technologie d’intégration par empilement 3D. Cette technologie permet d’intégrer un processeur dédié pour le traitement d’image dans le capteur. Un prototype du circuit RETINE a été réalisé en mettant en valeur la connexion pixels-processeurs dense dans le plan focal. La plupart des nouveaux concepts d’imageur couplant une matrice de pixels à des structures électroniques digitales gardent un goulot d’étranglement entre les couches pixels et numérique. L’imageur intelligent RETINE dépasse cette limitation habituelle de vitesse et bande passante grâce à la réplification de tuiles 3D. Cette tuile est composée d’une matrice 64x64 BSI pixels associée à un circuit de lecture de 16

ADC en pied de colonne sur la 1ère couche et couplée, dans le plan focal, à un processeur SIMD de 16 éléments de calcul sur la 2e couche. Cette technique permet, en plus de lever les limitations habituelles, une rétroaction rapide et localisée vers plusieurs zones du capteur d'image. En plus d'une meilleure connectivité entre les deux couches, ce concept est complètement scalable. Composé de 16x12 tuiles 3D, le prototype présenté peut appliquer des traitements d'image pendant la capture et ainsi, dans le même temps, ajuster les paramètres d'acquisition ou extraire des informations de l'image acquise. Il fonctionne à 80 MHz, pour une consommation de moins d'1 W.

14:55 **Camille Simon Chane**¹, **Sio Hoi Ieng**², **Ryad Benosman**^{2 3 4}

¹ *ETIS, UMR 8051, Université Paris Seine, Université Cergy-Pontoise, ENSEA, CNRS, FR*

² *Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau, 75012, Paris, FR.*

³ *University of Pittsburgh Medical Center, Biomedical Science Tower 3, Fifth Avenue, Pittsburgh, PA, USA.*

⁴ *Carnegie Mellon University, Robotics Institute, 5000 Forbes Avenue, Pittsburgh, PA, 15213-3890, USA.*

*Title: **Tone-mapping pour la visualisation de données issues de capteurs événementiels***

Abstract: Les caméras événementielles sont composées de pixels indépendants qui produisent une évènement lorsqu'il y a un changement d'intensité lumineuse dans leur champ de vue. Outre la polarité de ce changement (augmentation ou diminution d'intensité lumineuse) les caméras événementielles telles que l'ATIS mesurent la valeur de l'intensité. Le paradigme d'acquisition des caméras événementielles, basées donc sur un échantillonnage en intensité et non en fréquence temporelle, permet à l'ATIS d'imager des scènes avec une dynamique d'acquisition de 140 dB. Afficher une telle plage d'intensité lumineuse sur des écrans les 8-bits classiques soulève de nombreux problèmes : il est difficile de visualiser simultanément les zones sombres et lumineuses d'une scène, alors que ces données ont été acquises. Cela devient encore plus compliqué dans le cas d'une acquisition dont la luminosité globale change (scène intérieure/ extérieure par exemple). La compression de données à grande gamme dynamique pour l'affichage est dit "tone-mapping". Inspirée des solutions de tone-mapping image, les algorithmes présentés sont conçus de façon événementielle c'est à dire que pour chaque évènement entrant, on calcule uniquement la valeur d'affichage du pixel concerné. Une version globale de l'algorithme permet le traitement temps réel des données pour des taux d'évènements très élevés. Une version locale augmente le contraste locale et met donc en évidence des détails dans les zones sombres ou très éclairées. De plus nous proposons une version qui ajuste automatiquement les paramètres lorsque la luminosité globale de la scène change. Cet algorithme adaptatif tourne en temps réel sur un ordinateur portable pour des acquisitions allant jusqu'à 400 000 évènements par seconde.

16:30 **Kevin Hoang, Jean-Yves Dufour, Alexandre Pitti, Philippe Gaussier**
Thales GTS-SIX / ETIS Univ. de Cergy Pontoise - CNRS, FR

*Title: **Active exploration of a visual scene: impact of an event-based camera***

Abstract: In order to recognize one object, we process eye saccades movement to foveate its most relevant components.

We propose to model the visual scene recognition as an exploration process, where the evaluation confidence increase as the system actively gathers information in its visual field and to direct attention to other parts.

Our neural system combines the 'what' and 'where' information of 'what' has been seen and 'where' it has been seen (or 'how' it has been seen), learning to recognize an object and validating a sensorimotor approach as a way to build a robust representation for object recognition.

In a second experiment, we extend our framework to the use of a Prophesee event-based camera on a mobile platform to assess the validity of our approach.

We expect that the intrinsic dynamical properties of the camera will help pushing towards learning sensori-motor correlations since the data flow is only generated for a movement and provides an automatic segmentation for moving objects.

Furthermore, while preliminary experiments showed that it is possible to work with events integrated over time into frames, we will modify our neural model from a neural frequency based model to a spiking neural network to fully benefit from the event-based aspect of the camera. More precisely, we will make use of a Spike Timing-Dependent Plasticity (STDP) learning rule to synchronize the information delay of the various parts in our architecture.

POSTER LIST

Poster 1.**Stéphane Mancini***Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMA, Grenoble, FR**Title: **Caméléon: un prototype de caméra pan-optique sur FPGA*****Poster 2.****Victor Boutin**^{1,2}, **Angelo Franciosini**¹, **Franck Ruffier**², **Laurent Perrinet**¹¹ *Aix Marseille Univ - CNRS, INT, Marseille, FR*² *Aix Marseille Univ - CNRS, ISM, Marseille, FR**Title: **Sparse Deep Predictive Coding To Model Visual Object Recognition***

Convolutional Neural Networks (CNNs) are the state-of-the-art algorithms used in computer vision. However, these models often suffer from the lack of interpretability of their information transformation process. To address this problem, we introduce a novel model called Sparse Deep Predictive Coding (SDPC). In a biologically realistic manner, SDPC mimics how the brain is efficiently representing visual information. This model complements the hierarchical convolutional layers found in CNNs with the feed-forward and feed-back update scheme described in the Predictive Coding (PC) theory and found in the architecture of the mammalian visual system. We experimentally demonstrate on two databases that the SDPC model extracts qualitatively meaningful features. These features, besides being similar to some of the biological Receptive Fields of the visual cortex, also represent hierarchically independent components of the image that are crucial to describe it in a generic manner. For the first time, the SDPC model demonstrates a meaningful representation of features within the hierarchical generative model and of the decision-making process leading to a specific prediction. A quantitative analysis reveals that the features extracted by the SDPC model encode the input image into a representation that is both easily classifiable and robust to noise.

Poster 3.**Tushar Chauhan, Timothée Masquelier, Paul Fricker, Benoit R. Cottureau***CerCo, Université de Toulouse, Cnrs, UPS, Toulouse, FR**Title: **A bio-inspired STDP network for event-based unsupervised learning***

Abstract: Biological vision has evolved in response to our environment for millions of years. It is characterised by highly optimised structures which compress and process information at various stages of the visual system from sensing to perception. Recent years have seen the emergence of bio-inspired image acquisition technologies which leverage biological optimisations towards the development of faster, more energy efficient systems. Most of these systems are based on asynchronous or near-asynchronous event-based sampling of the visual environment.

In this talk, we present an unsupervised event-based network with spike-timing dependent plasticity (STDP). STDP is a biologically reported mechanism of Hebbian plasticity which relies on the latencies of input signals. We show how our network not only converges to realistic neural populations, but that when interfaced with real event-based cameras, it can lead to meaningful data-compression which can be used to perform tasks such as pattern-recognition. Such data-compression could be highly useful in situations where learning is not goal-directed, but an initial step towards the learning of multiple tasks using a single set of training events.

Poster 4.

Veis Oudjail, Jean Martinet

CRISTAL, Univ. Lille - CNRS, Lille, FR

Title: Étude du mouvement d'un motif à l'aide d'un réseau de neurones impulsionnels

Abstract: Pouvoir estimer la direction du mouvement est une tâche essentielle en analyse vidéo. Le flux optique quantifie le mouvement apparent des objets. L'estimation du mouvement peut se faire grâce à des modèles s'inspirant du fonctionnement biologique de la rétine des mammifères, où des cellules ganglionnaires montrent des dispositions à percevoir des directions globales, en particulier sur les quatre directions cardinales. Cette approche contrebalance les méthodes basées sur le deep-learning généralement employées, qui s'illustrent, certes, avec des résultats supérieurs aux autres méthodes, mais dont l'apprentissage nécessite une grande quantité de données ainsi que des ressources de calculs importantes. C'est pourquoi nous proposons un nouveau modèle d'estimation du mouvement, basé sur des caméras event-based (DVS par exemple), qui sont des capteurs bio-inspirés sensibles au contraste temporel d'une scène, couplé à un réseau de neurones impulsionnels (Spiking Neural Network, SNN) de taille minimale, qui apprend les directions du mouvement de manière non-supervisée. Les SNN sont une classe particulière de réseau de neurones, proche du modèle biologique, dans lequel les neurones émettent des impulsions sortantes (spikes) de manière asynchrone, en fonction des stimulations entrantes, asynchrones elles aussi. L'apprentissage avec de tels réseaux repose sur une règle "Spike-Timing Dependent Plasticity", qui met à jour les poids synaptiques au cours d'une simulation en fonction des relations de cause à effet constatées entre les impulsions entrantes et sortantes. Le but de cette règle est le renforcement des connexions entrantes qui sont la cause des impulsions sortantes. Les avantages des SNN sont les capacités d'apprentissage non supervisé et continu, ainsi que leur possible mise en œuvre sur du matériel dédié à très faible consommation énergétique et massivement parallèle du fait de la localité de la règle d'apprentissage. Après quelques millisecondes de simulation, les détecteurs de mouvement proposés apprennent automatiquement à détecter les directions de mouvement. Le modèle est entraîné à l'aide d'un ensemble de données événementielles synthétiques composé de motifs aléatoires se déplaçant dans plusieurs directions. Les résultats expérimentaux montrent que de tels réseaux peuvent apprendre et estimer avec succès les différentes directions de manière non-supervisée.