



HAL
open science

Enjeux de la fovéation appliquée aux données événementielles

Amélie Gruel, Jean Martinet, Bernabé Linares-Barranco, Teresa Serrano-Gotarredona

► **To cite this version:**

Amélie Gruel, Jean Martinet, Bernabé Linares-Barranco, Teresa Serrano-Gotarredona. Enjeux de la fovéation appliquée aux données événementielles. ORASIS 2021, Centre National de la Recherche Scientifique [CNRS], Sep 2021, Saint Ferréol, France. hal-03339647

HAL Id: hal-03339647

<https://hal.science/hal-03339647>

Submitted on 9 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enjeux de la fovéation appliquée aux caméras événementielles

Stakes of foveation on event cameras *

Amélie Gruel¹ Jean Martinet¹ Bernabé Linares-Barranco² Teresa Serrano-Gotarredona²
¹*Université Côte d’Azur, CNRS, I3S, Sophia-Antipolis, France*
²*Instituto de Microelectrónica de Sevilla IMSE-CNM, Sevilla, Spain*
amelie.gruel@univ-cotedazur.fr

Résumé

La fovéation consiste en la direction du regard vers une région visuelle d’intérêt pour acquérir sélectivement des informations pertinentes. Dans l’avènement récent des caméras événementielles, nous pensons que le développement d’un tel mécanisme améliorerait l’efficacité du traitement des données événementielles. En effet, appliquer la fovéation à celles-ci permettrait de comprendre la scène visuelle tout en réduisant considérablement la quantité de données brutes à traiter. Nous étudions l’évolution de la précision de la segmentation par rapport à la quantité de données utilisées, pour démontrer les enjeux de la fovéation.

Mots clés

Fovéation, caméras événementielles, segmentation, saillance.

Abstract

Foveation is the organic action of directing the gaze towards a visual region of interest, to selectively acquire relevant information. In the recent advent of event cameras, we believe that developing such a mechanism would greatly improve the efficiency of event-data processing. Indeed, applying foveation to event-based data would allow to comprehend the visual scene while significantly reducing the amount of raw data to handle. We study the evolution of the accuracy of segmentation with respect to the amount of event data used, to demonstrate the stakes of foveation.

Keywords

Foveation, event cameras, segmentation, saliency.

1 Introduction

Event cameras (or silicon retinas) represent a new kind of sensors that measure pixel-wise changes in brightness and output asynchronous events accordingly [1]. This novel technology allows for an energy-efficient recording and storage of data evolving over time and

space. Indeed each event is recorded punctually and asynchronously with no redundancy (as opposed to traditional cameras, where each pixel records information at all time).

Foveation is the biological action allowed by the structure of the complex eye [2]. When the gaze is directed towards a region of interest, the center of the perceptual field is caught by the fovea, a small but central spot in the retina where the vision is optimal in bright light. The further we get from the fovea, the lesser the information is processed by the eye. This mechanism is driven by the visual attention [3].

We believe that developing a mechanism approaching foveation would greatly improve event data processing, going beyond its significant energy-efficiency. Indeed, this would allow to maintain a high accuracy regarding relevant information, while significantly reducing the amount of raw data to handle. Furthermore, this approach consistency is supported by the fact that silicon retinas aim by definition to reproduce the biological retina behaviour.

2 Proposed approach

Fig. 1 presents succinctly the proposed approach as proof of concept. To demonstrate the interest in applying foveation to event-based data, we propose to study the respective evolution of the amount of event data processed by a segmentation algorithm and its accuracy. The event data will be processed at a higher

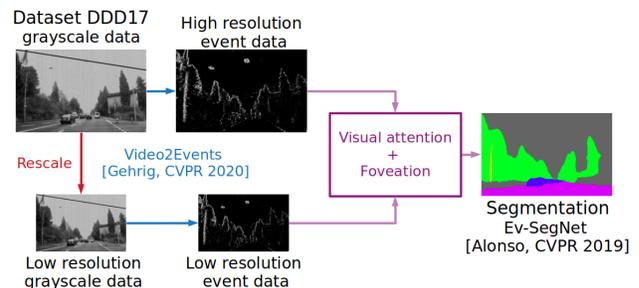


FIGURE 1 – Proposed approach to demonstrate the stakes of foveation in event data processing.

*This work was supported by the European Union’s ERANET CHIST-ERA 2018 research and innovation programme under grant agreement ANR-19-CHR3-0008.

or lower resolution, depending on the relevance of the spatial regions in the image at different coordinates. This approach is part of the work conducted in the context of APROVIS3D¹ project, and is ultimately to be applied to the use case of coastline tracking by a UAV. Thus the dataset has to be chosen in order to approach this use case, as was the segmentation task.

2.1 Input data

Given a foveation simulator, the event-based data will be acquired either at a higher or a lower resolution from a grayscale dataset, or any combination in a dedicated coordinate space.

Dataset The *DAVIS Driving Dataset 2017 DDD17* [4] provides the recording of a highway and city driving in different driving conditions under the form of events and grayscale videos. This dataset is particularly of interest since it comes near the kind of visual data that will be used by the APROVIS3D consortium.

High to low resolution The high-resolution flow of events is 128×128 pixels in area, and the low-resolution flow is 64×64 pixels. The dedicated coordinate space used to define regions of interest and gather high or low-resolution data corresponds to the high-resolution's, with a mapping function for low-resolution. The low resolution is acquired by rescaling the high resolution grayscale images.

Traditional videos to events The traditional grayscale video data at high and low resolution needs to be translated into events. We consider Gehrig and Scaramuzza's *Video to events Vid2e* [5].

2.2 Segmentation algorithm

We propose to use as baseline the segmentation algorithm *EV-SegNET* presented in CVPR 2019 by Alonso and Murillo [6]. The authors selected from *DDD17* six sequences of grayscale images from urban scenarios and annotated them in order to train and test the model. The model itself consists in an encoder-decoder architecture for semantic segmentation on event images.

2.3 Detection of regions of interest

The detection of regions of interest to foveate on is a little explored issue regarding event-data. We propose to test three mechanisms in increasing complexity.

Event density A first possibility could be to implement a novel spiking neural network, integrating the events produced by each pixel at a low resolution and outputting a set of coordinates to the region of interest. In this case, the region of interest would be the region where the amount of events received over a certain amount of time is the most important over the whole scene.

Saliency Another way to detect region of interests could be to adapt Itti's attention model in multimedia to event data [7]. This model detects saliency according to features such as intensity, orientatin, etc. to achieve rapid scene analysis. This detector has been noticeably implemented in robotics applied to event cameras by Bartolozzi's team [8].

Gestalt laws Lastly we could draw inspiration from another work of Bartolozzi's team on proto-objects [9]. The regions of interest in the visual scene could indeed match those proto-objects, derived from the "Gestalt laws" [10] and defined as regions of the scene that potentially corresponds to physical objects.

3 Conclusion

We aim to demonstrate the stakes of foveation applied to event-based data with this proposed approach. We believe that such a strategy would concurrently preserve the accuracy of event-data processing and greatly improve the energy-efficiency of event data.

Références

- [1] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128x128 120 db 15 ms latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2) :566–576, 2008.
- [2] M.F. Bear, B.W. Connors, and M.A. Paradiso. The Human Eye. In *Neurosciences, Exploring the brain*, Wolters Kluwer Health. 2007.
- [3] Amélie Gruel and Jean Martinet. Bio-inspired visual attention for silicon retinas based on spiking neural networks applied to pattern classification. *arXiv [cs]*, 2021.
- [4] J. Binas, D. Neil, S.-C. Liu, and T. Delbruck. DDD17 : End-To-End DAVIS Driving Dataset. *arXiv :1711.01458 [cs]*, 2017.
- [5] D. Gehrig, M. Gehrig, J. Hidalgo-Carrio, and D. Scaramuzza. Video to Events : Recycling Video Datasets for Event Cameras. In *CVPR*, 2020.
- [6] I. Alonso and A.C. Murillo. EV-SegNet : Semantic Segmentation for Event-based Cameras. *CVPR Workshops*, 2019.
- [7] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 20(11), 1998.
- [8] F. Rea, G. Metta, and C. Bartolozzi. Event-driven visual attention for the humanoid robot iCub. *Frontiers in Neuroscience*, 7, 2013.
- [9] M. Iacono, G. D'Angelo, A. Glover, V. Tikhanoff, E. Niebur, and C. Bartolozzi. Proto-object based saliency for event-driven cameras. In *IROS*, 2019.
- [10] W. Köhler. Gestalt psychology. *Psychologische Forschung*, 31(1) :XVIII–XXX, 1967.

1. URL : www.aprovis3d.eu