Multimodal perception of vulnerable road users for autonomous driving in urban environments

Vincent BREBION
Directeurs de thèse : Franck DAVOINE, Julien MOREAU
Équipe : SyRI

***Abstract*—Urban areas are complex driving environments where vehicles and personal mobility devices (pedestrians, cyclists, ...) often operate close to each other. Autonomous driving in these environments represents a challenge, due to the highly unpredictable behaviour of these vulnerable users. The objective of the thesis is therefore to develop multimodal perception systems to improve their detection and avoid any collision by means of computer vision, multi-sensor fusion, and machine learning.**

I. INTRODUCTION

Autonomous driving in open, uncontrolled environments calls for deep understanding abilities from the self-driving vehicle, to make it able to navigate safely. This understanding process requires the detection and recognition of potential obstacles. The use of perception sensors allows for such capabilities, especially powered by the recent rise of machine-learning-based methods, reaching never-achieved-before heights in recognition tasks.

Most of the results from the literature, however, were achieved in favorable conditions (adequate lighting and weather, clearly visible objects), which only represent a fraction of the real-life situations a driver is confronted with. Recent studies have particularly shown the limits of these approaches in more complex conditions (at dawn/dusk, during rain/snow, when the vulnerable user is partially occluded by another object or very close to the ego-vehicle, ...), raising multiple safety questions [1], [2].

In parallel, the navigation in urban environments has been deeply changing over the past few years, with the rise of soft mobility solutions (bicycles, scooters, skateboards, rollerblades, hoverboards, ...). While they allow for more flexible movements in urban areas, they especially put their user at risk in case of a collision with a traditional vehicle. This risk is further amplified by the erratic behaviour these users may have: slaloming between cars, alternating between the use of the road and the sidewalks, not respecting the road markings, navigating close to the other vehicles, etc.

As an answer to these issues, the objective of the thesis is to reinforce the detection of these vulnerable users in difficult visual conditions typical in urban environments. To reach it, two complementary approaches are going to be used:

- the arrival of novel sensors (e.g. event-based cameras) and the improvement of others (e.g. lidars, thermal cameras) open new perception capacities, even in complex lighting and/or weather conditions; using them for detection tasks in such situations therefore appears as an interesting baseline to explore;
- the use of multi-sensor fusion techniques, allowing for safety, redundancy, and better and more stable detection results, also appears as an important component, and an environment-aware framework should be employed to dynamically select the best sensors to use based on the current lighting and weather conditions.

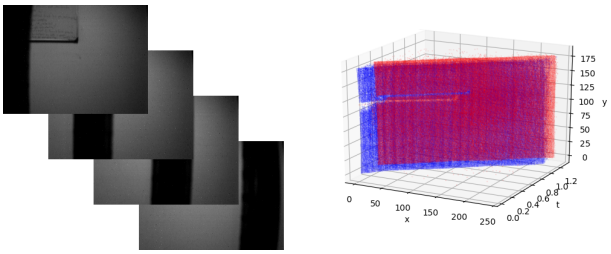

Fig. 1. Traditional images (left) vs events (right), for a simple "pen moving from left to right in front of the camera" sequence. Events are ploted in 3D along the x, y, and t axes, and each dot represents an event (blue ones denote an increase of lighting for the corresponding pixel at the given time, red ones a decrease)

II. INITIAL ORIENTATION OF THE THESIS

While event-based sensors do not represent the only central component that are planned to be used for improving the detection of vulnerable users, their recency and their ever-growing rise in popularity makes them highly attractive in the field of computer vision as of today. Furthermore, while these sensors were limited to low-resolutions in the past, new high-definition ones have started to emerge in 2020 [3], opening new usage perspectives. As such, and thanks to the collaboration with Prophesee, which are one of the main manufacturers of event-based sensors, we were given the opportunity to have access to one of their high-resolution prototypes, and thus, the chance of being among the first to publish with such cameras arose.

The focus for this first year of PhD thesis was therefore set on these cameras, and in particular on their use for computing optical flow. This issue indeed constitutes a central problem in the field of computer vision, and it is a key enabler for other applications, including especially the object detection and recognition issues.

III. EVENT-BASED OPTICAL FLOW

In opposition to "conventional" frame-based cameras, which accumulate light during short periods of time to create dense

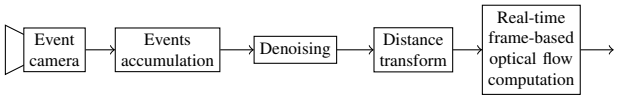

Fig. 2. Our optical flow computation architecture, able to run in real-time with low- and high-resolution event cameras. Blocks are run in parallel.

images, event cameras produce only small packets of information (called events), which report asynchronously and independently per pixel the changes of light in the observed visual scene (see Fig. 1). By doing so, they represent a great paradigm shift in the field of computer vision, opening new ways of sensing the world. This shift, however, implies that state-of-the-art computer vision methods developed over the past decades for traditional cameras can not be directly employed with event-based sensors.

For that reason, two main visions on how to approach events can be distinguished in the literature. On one hand, multiple researchers argue that, due to their highly specific characteristics, events should be used as is, and that new methods for solving computer vision issues should be developed. On the other hand, many authors point out the already existing literature for approaches using frame-based sensors, and that being able to reuse them by recreating frame-like structures from events could allow for the use of these cameras in real-life applications in a near future.

Our goal being a use with autonomous vehicles, we therefore settled for the second option, which appeared as more realistic for a short-term real-life application, and allowed us to reuse a state-of-the-art method for computing optical flow.

Thus, as part of the thesis, we developed a real-time framework, allowing for the computation of optical flow with both low- and high-resolution event-based sensors. This framework uses a pipeline architecture, composed of four main blocks, and is presented in Fig. 2. The first three blocks are respectively tasked with event accumulation over short temporal windows, denoising, and smart densification, allowing for final dense "images" (specifically designed to be used for optical flow computation) to be created from the events. These images are then used as the input to a proven, state-of-the-art real-time frame-based optical flow library [4], allowing us to obtain our final optical flow results.

While the real-time constraint limited us in the complexity of the pipeline, we still were able to achieve optical flow results close to the non-real-time state of the art for low-definition event cameras, and were also able to show the correctness of our results for complex high-definition recordings. Furthermore, our framework is capable of frame rates of 250Hz and 77Hz for resolutions of 346×260 and 1280×720 respectively, thus enabling its use in a real-life context. We present in Fig. 3 visual results from a high-definition urban sequence. More complete video results for various sequences are also available if the reader is interested: see https://youtube.com/playlist?list=PLLL0eWAd6OXBRXli-tB1NREdhBElAxisD.

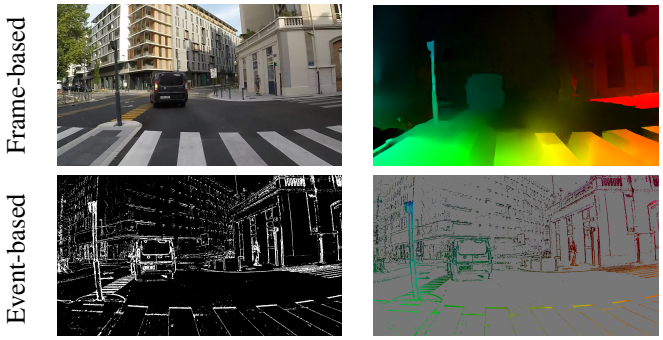

Fig. 3. Optical flow results, for a urban sequence (dataset kindly provided by Prophesee). Top row: reference RGB image, and optical flow reference computed from this image using a state-of-the-art optical flow network, RAFT [5]. Bottom row: events received from the event-based sensor and accumulated over a short temporal window (white pixels), and our optical flow results from them (zones in gray represent pixels for which no event was received, and therefore no flow was computed). Notice especially how our optical flow results are visually close to the frame-based reference.

IV. Publication

As part of this work, a publication titled "Real-Time Optical Flow for Low- and High-Resolution Event Cameras" is being prepared, and is planned to be submitted to IEEE Transactions on Intelligent Transportation Systems before the end of July 2021. Our own datasets involving high-definition event camera are also planned to be made available on Heudiasyc's platform (https://datasets.hds.utc.fr).

V. Conclusion and Future Work

As of today, this work on real-time optical flow computation using event-based cameras is near completion. Once finished, the focus will be set to bibliographic work, to determine the future orientation of the thesis, and the work which will be carried out in the next two years.

Acknowledgment

This work is supported by the Hauts-de-France Region (through the ALRC program) and SIVALab (Renault-UTC-CNRS).

References

[1] B. A. T. Brown and E. Laurier, "The trouble with autopilots: Assisted and autonomous driving on the social road," *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017.

[2] T. S. Combs, L. S. Sandt, M. P. Clamann, and N. McDonald, "Automated vehicles and pedestrian safety: Exploring the promise and limits of pedestrian detection." *American journal of preventive medicine*, vol. 56 1, pp. 1–7, 2019.

[3] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, É. Reynaud, P. Mostafalu, F. T. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike, and C. Posch, "A 1280×720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86µm pixels, 1.066GEPS readout, programmable event-rate controller and compressive data-formatting pipeline," *IEEE International Solid- State Circuits Conference (ISSCC)*, pp. 112–114, 2020.

[4] J. Adarve and R. Mahony, "A filter formulation for computing real time optical flow," *IEEE Robotics and Automation Letters*, vol. 1, pp. 1192–1199, 2016.

[5] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *ECCV*, 2020, pp. 402–419.