Reg. No: 2024/53/N/ST6/04254; Principal Investigator: mgr inż. Kamil Andrzej Jeziorek

F+E: Enchancing Perception through the Integration of Frame and Event Cameras

In recent years, a growing interest in a new type of vision sensor, called an event camera, can be seen in both the scientific and industrial communities. Their popularity is due to their unique properties, inspired by the structure of the human eye, which distinguish them from the classic frame-based cameras familiar from our mobile phones.

Both types of sensors consist of a matrix of pixels recording light intensity. Traditional so-called frame cameras read information from all pixels simultaneously at regular intervals (e.g. every 20 ms), returning an image in the form of a photo or video frame. In contrast, event cameras work differently: each pixel operates independently and asynchronously, monitoring changes in light intensity. When the change exceeds a set threshold, the pixel generates a single piece of information called an event. Figure 1 shows the outputs from both types of camera, illustrating the differences between them. In the case of an event camera, only information is recorded where a change has occurred.

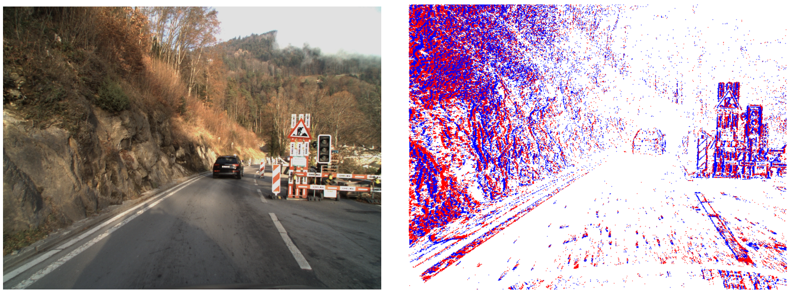

Figure1. Comparison of results from a frame-based and event-based camera. The red and blue colours indicate the positive or negative polarity of the illumination change for a given pixel, respectively.

Of course, due to the different operating principle, both technologies have their advantages and disadvantages. In the case of event cameras, the advantage is that they are more resistant to the so-called image blur effect and work relatively better in more dynamic situations. Unfortunately, they work less well in more static scenarios because they generate much less information. Frame cameras, on the other hand, provide continuous information even in the absence of movement. Combining the two technologies can create an effective vision system, used, for example, in the safety systems of modern cars.

Recent work on vision systems has focused on integrating the two types of cameras, which shows improvements in their performance. Most of these solutions are based on artificial intelligence, analysing data from both sources. However, most of the research carries out its work using GPU computing platforms, which are not suitable for autonomous drones or car safety systems due to their large size and high power consumption. Better in these cases are embedded computing platforms such as eGPUs (*embedded GPU*s) or SoC FPGAs (*System-on-Chip Field-Programmable Gate Arra*y), which are lighter and consume less power, but have limited computational capabilities, which requires adaptation of models.

Therefore, the project seeks to investigate existing solutions integrating both types of sensors and their implementation on embedded systems. The aim is to assess the applicability of these solutions on such platforms and to analyse the potential losses compared to the original systems. In addition, it is planned to develop proprietary models and algorithms with hardware acceleration in mind from the beginning. As part of data integration, it is planned to apply the latest techniques used in artificial intelligence such as knowledge-distillation between models, attentional mechanisms, state space models, and techniques used in multi-modal models.

The work will result in a set of algorithms and methods tailored to the limitations of embedded computing platforms, applicable to autonomous vehicles such as cars and drones. Implementation of these models will improve their efficiency and reliability, enhancing safety. The work will also provide a better understanding of the advantages and disadvantages of different methods for simplifying data-integrating models, which will be useful for the design of future solutions in this field.