Sustainable event-based vision for scene perception and understanding by autonomous robots

Current neural network models have enabled significant advancements in object detection, event recognition, and the identification of human behaviors and actions-achievements that would have seemed nearly impossible just a decade ago. This progress inspires us to explore future possibilities and develop scientific hypotheses about how intelligent robots might achieve a deeper understanding of events and image content. Given the rapid technological advances in robotics and artificial intelligence in recent years, many leading experts predict that, within the next 20 to 30 years, the number of humanoid robots could surpass the human population. In the near future, we anticipate a substantial rise in the deployment of autonomous robots across homes, hospitals, and offices, leading to transformative changes in the quality of life and the way humans live and work. However, despite these advancements, modern mobile robots still face significant challenges in meeting the critical requirements for safe and efficient operation in environments shared with humans. These challenges are partly due to the perceptual limitations of the cameras in current robotic vision systems, as well as the high energy consumption associated with the way traditional neural networks process sequences of images, which can quickly deplete onboard battery power.

In robotic vision systems, the effectiveness of scene perception is a critical factor that determines the overall performance of the system. However, without a fundamental shift in how dynamic scenes and the robot's environment are represented–particularly through significant advancements in decision-making methods that operate autonomously, based on real-time sensor data–it will be challenging to scale these systems effectively. A key limitation of current object detection and image recognition techniques is their high energy consumption, which is orders of magnitude greater than that required by humans, both during the learning phase and when performing routine tasks.

There is currently a growing demand for computationally efficient and environmentally sustainable methods of perception and scene understanding for autonomous systems. Specifically, there is a need for more advanced and effective solutions for perceiving dynamic environments, particularly those involving human interactions. The research in this project will focus on the development of neural networks that process raw data from neuromorphic sensors, also known as event-based sensors. Unlike traditional vision sensors, which generate continuous and dense data, event cameras capture data asynchronously and sparsely, presenting a fundamentally different paradigm for processing visual information.

The research will focus on developing the theoretical framework for neural networks that can adjust their time-scale parameters in real-time, based on the content, complexity, and dynamics of scenes and events. It will also explore selective approaches for processing sparse event streams. In this context, state-space methods from control theory and self-attention mechanisms will be leveraged. The second research pathway will involve the development of competitive solutions for hybrid event data and image maps, including innovative object tracking methods that effectively mitigate the issue of losing an object's trajectory by utilizing event streams during inter-frame intervals. The third focus will be on developing competitive solutions for human-object-machine interactions, with an emphasis on motion analysis based on event data captured by dynamic onboard cameras. The project outcomes will include not only solutions addressing key open questions in computer/robot vision but also practically validated algorithms and specific solutions for sustainable robot vision, as well as dynamic scene and environment perception, using real-world data from dynamic cameras.