Human-centric vision with event-cameras: Human Pose Estimation

- Gaurvi Goyal, Istituto Italiano di Technologia

Event-based cameras are gaining popularity recently, being ideal candidates for many vision tasks due to their unique features. Among these, since they capture only the change in the scene, they support development of light-weight and asynchronous processing modules, ideal for systems with constraints on latency allowance, energy or computation power, like mobile robots or smart phones.

Human-machine interaction is a cornerstone for such systems. Accurately detecting the State of the person can be crucial to smooth interaction. State can be defined by any relevant factors regarding the mental and physical state of a person. These can be, but are not limited to, body pose, action being performed, human-object interactions, eye gaze direction, emotional state, fatigue level and cognitive load level.

As of now, our focus is on Human Pose Estimation (HPE) with event cameras, tracking 13 keypoints on the visible person. We developed methods to leverage latest development in Computer Vision and Machine Learning, and publicly available image and video datasets, bootstrapping a lightweight Convolutional Neural Network (CNN) for HPE with event cameras, MoveEnet. In addition, we created a model based tracker for the joints, and together they can estimate human pose accurately on a CPU at upto 500Hz, directly from the camera.

CNNs, are widely used and effective, but are designed for dense inputs and synchronous, forgoing the unique properties of an event camera. Therefore, we are currently developing a Graph Neural Network (GNN), more appropriate for the sparse event data, while still being lightweight and within the constraints outlined earlier. Though there are recent works towards GNNs for event inputs, but as of writing this, we are unaware of any GNNs for human-centric applications from event cameras, or any light-weight GNNs for event camera applications.. We have promising preliminary results from this system, although more time is needed for create an accurate, precise, stable and robust system.