Learning from actions - Temporal structures for human action recognition

Action recognition has become a more and more important topic in academic as well as in industrial research. So far, most approaches in the field were focusing on the classification of hand selected preclipped actions. But to address the inherent complexity of large scale data, video-based action recognition is more and more advancing from simple clip classification towards the analysis of longer video sequences. But it shows that this step not only depends on the development of better feature descriptors and learning architectures, but also on the handling of the inherent temporal information that comes with this type of data.

This talk focuses on the analysis of video sequences by modeling activities as temporally structured processes. As temporal structures are a key element for our understanding of video sequences, a hierarchical architecture similar to human action perception is proposed. It allows to capture temporal information at different levels of granularity, ranging from short movements to task oriented activities with high variation. Experiments and use cases show that the proposed temporal modeling and segmentation not only leads to a better recognition and parsing of videos, it also allows to go beyond the usage of fully annotated training data and address the problem of weak and unsupervised learning of human actions. The proposed concepts will be a first step beyond the necessity of predefined, pre-clipped actions for training, towards a less restricted recognition and analysis of human actions.