

Dickmanns 2015: **Knowledge bases for visual dynamic scene understanding**

**Abstract:** In conventional computer vision the actual 3-D state of objects is of primary interest; it is embedded in a temporal sequence analyzed in consecutive pairs. In contrast, in the 4-D approach to machine vision the primary interest is in temporal processes with objects and subjects (defined as objects with the capability of sensing and acting). All perception of 4-D processes is achieved through feedback of prediction errors according to spatiotemporal dynamical models constraining evolution over time. Early jumps to object/subject-hypotheses including capabilities of acting embed the challenge of dynamic scene understanding into a richer environment, especially when competing alternatives are pursued in parallel from beginning. Typical action sequences (maneuvers) form an essential part of the knowledge base of subjects. Expectation-based Multi-focal Saccadic (EMS-) vision has been developed in the late 1990s to demonstrate the advantages and flexibility of this approach. Based on this experience, the paper advocates knowledge elements integrating action processes of subjects as general elements for perception and control of temporal changes, dubbed 'maneuvers' here. – As recently discussed in philosophy, emphasizing individual subjects and temporal processes may avoid the separation into a material and a mental world; EMS-vision quite naturally leads to such a monistic view.