

# Describing When and Where in Vision<sup>\*</sup>

Walter G. Kropatsch, Adrian Ion, and Nicole M. Artner

PRIP, Vienna University of Technology, Austria

[krw@prip.tuwien.ac.at](mailto:krw@prip.tuwien.ac.at)

<http://www.prip.tuwien.ac.at/>

**Abstract.** Different from the what and where pathways in the organization of the visual system, we address representations that describe dynamic visual events in a unified way.

Representations are an essential tool for any kind of process that operates on data, as they provide a language to describe, store and retrieve that data. They define the possible properties and aspects that are stored, and govern the levels of abstraction at which the respective properties are described. In the case of visual computing (computer vision, image processing), a representation is used to describe information obtained from visual input (e.g. an image or image sequence and the objects it may contain) as well as related prior knowledge (experience).

The ultimate goal, to make applications of visual computing be part of our daily life, requires that vision systems operate reliably, nearly anytime and anywhere. Therefore, the research community aims to solve increasingly more complex scenarios. Vision both in humans and computers is a dynamic process, thus variations (change) always appear in the spatial and the temporal dimensions. Nowadays significant research efforts are undertaken to represent variable shape and appearance, however, joint representation and processing of spatial and temporal domains is not a well-investigated topic yet. Visual computing tasks are mostly solved by a two-stage approach of frame-based processing and subsequent temporal processing. Unfortunately, this approach reaches its limits in scenes with high complexity or difficult tasks e.g. action recognition. Therefore, we focus our research on representations which jointly describe information in space and time and allow to process data of space-time volumes (several consecutive frames).

In this keynote we relate our own experience and motivations, to the current state of the art of representations of shape, of appearance, of structure, and of motion. Challenges for such representations are in applications like multiple object tracking, tracking non-rigid objects and human action recognition.

---

<sup>\*</sup> Supported by the Austrian Science Fund under grants P20134-N13 and P18716-N13.