What happens where in video?

Cees Snoek
University of Amsterdam & Qualcomm Research Netherlands

Abstract

This lecture is about learning to see in video, a branch of computer vision that has witnessed considerable progress during the past decade. The major challenge in video recognition is to automatically understand what is happening where in the visual content. The lecture will first highlight traditional invariant representations and supervised deep learning algorithms to detect concepts such as 'person', 'boat' and 'beach' in video-fragments. We do so by reviewing the winning solutions in the TRECVID benchmark, the leading video search engine competition organized yearly by the US National Institute of Standards and Technology. After establishing the basics, the lecture will dive into two emerging areas of investigation: recognizing actions and understanding events in web video.

The leading action recognition techniques emphasize on encoding motion characteristics to cover the many fine-grained spatiotemporal variations in human action appearance. In addition, motion is a valuable cue for localization of actions such as 'boxing' and 'hand waving'. The lecture will detail recent algorithms that automatically segment video sub-volumes encompassing an action of interest. In addition, we will share the findings of an empirical study on the benefits of encoding thousands of concept categories for action classification and localization.

In the last part the lecture will zoom in on events: descriptions of video content combining concepts and actions into sentences, like 'working on a wood working project' and 'winning a race without a vehicle'. Initially the detection of events followed the same recognition conventions as concepts and actions, but recently this regime has been challenged by semantic embeddings. We will detail semantic

embeddings that allow for accurate detection and are also able to translate and summarize events in video content, even in absence of training examples.
The lecture concludes with a perspective on future challenges and opportunities for video recognition.

Keywords

Concepts, actions, events

Speaker

Cees Snoek is an associate professor at the University of Amsterdam and principal engineer at Qualcomm Research Netherlands. He was previously visiting scientist at Carnegie Mellon University, Fulbright scholar at UC Berkeley and head of R&D at Euvision Technologies (acquired by Qualcomm). His research interests focus on video and image retrieval. Dr. Snoek is the lead researcher of the MediaMill team, which is the most consistent top performer in the yearly NIST TRECVID video search evaluations. Cees is recipient of several best paper and career awards, including the Netherlands Prize for ICT Research. He is general co-chair of ACM Multimedia 2016 in Amsterdam.