

Multimodal Learning and Computing for Human Activity Understanding

Introduction

Automatically understanding and recognizing human activity from video is one of the most active research areas in the multimedia community due to the scientifically challenging problems and its great benefits to real life applications. Its development brings contributions to fields such as surveillance systems, patient monitoring systems, elderly care, intelligent robots, athletic training and human-computer interfaces (HCI). Multiple modalities of information are involved in video streams, which characterize the nature and properties of the underlying human activity events. For instance, visual cues may include people, objects and their actions, audio cues can show characteristic sounds in corresponding actions, while dialogues, text floating around video key frames may be deemed as textual events. By combining these modalities, visual recognition systems can achieve more promising results due to the richer nature of the data.

This special session calls for submissions on the emerging and challenging research topics on human activity understanding and learning by exploiting the different modalities in video and integrating the information in a non-trivial way.

- The topics of interests include, but are not limited to, the following:
- Multimodal representation and feature descriptors, feature selection and fusion for human activity understanding
- Multimodal approaches to detecting actions and complex activities
- Multimodal approaches to event analysis and modeling for human activity understanding
- Machine learning for multimodal analysis
- Real-world event/trend detection, localization, tracking and visualization from social multimedia
- Applications and services on human activity understanding