Tutorial 14: Multisensory Video Processing and Learning for Human-Robot Interaction

In many human-robot interaction (HRI) application areas where multisensory processing and recognition is greatly needed, multimodality occurs naturally and cross-modal integration increases performance. This tutorial addresses the multisensory spatio-temporal processing of visual information together with its fusion with the speech/audio modality as applied to two different HRI areas: assistive and social robotics. Our coverage will include theory, algorithms and a rich variety of integrated applications for specific groups like elderly users and children. There are many challenges in this area including the familiarity of these users with new technologies and the domain specific datasets, which are required for training user oriented models. Nowadays, modern assistive and social HRI requires a multimodal communication with speech, gestures and human movements so as to enhance the classic interaction with only spoken commands. This tutorial will present state-of-the-art works for multisensory and visual processing and machine learning models that can be effectively trained with a relatively small amount of data, which is very important when we deal with elderly users and children. Moreover, in the present state of our information society, we are witnessing a very rapid expansion of multimodal and multisensory content, with huge volumes of multimedia content being continuously created. As a result, multimodal processing technologies have become increasingly relevant. Computer vision techniques, despite recent advances, still significantly lag behind the human ability in understanding real-life scenes and performing demanding robotic tasks. Motivated by the multimodal way humans perceive their environment, complementary information sources have been successfully used in many applications, such as human action recognition where the audio-visual cues pose many challenges at the level of features, information stream modeling and fusion. Afterwards, we will focus on the major application area, which is Human-Robot Interaction, for social, edutainment and healthcare applications, including audio-gestural commands recognition and multi-view human action recognition. Established results and recent advances from our research in the three EU projects MOBOT, I-SUPPORT and BABYROBOT will also be discussed. More information and related papers can be found in http://cvsp.cs.ntua.gr and http://robotics.ntua.gr.

Speakers:

Petros Maragos
National Technical University of Athens, School of ECE, Athens, Greece.
Athena Research Center, Robot Perception and Interaction Unit, Greece.

Petros Koutras
National Technical University of Athens, School of ECE, Athens, Greece.
Athena Research Center, Robot Perception and Interaction Unit, Greece