Praxis der Forschung

Tamim Asfour

Institute for Anthropomatics and Robotics (IAR)
High Performance Humanoid Technologies (H^{2}T)

Deep Learning for Semantic Scene and Action Perception

When a robot needs to imitate an observed action sequence, it must first understand the inherent characteristic features of individual actions that remain the same even under different circumstances. These descriptive features reflect the semantics of actions which has high degree of invariance between different demonstrations of the same action in various scene contexts. In this sense, the more descriptive and structural the action semantics, the greater the capabilities and autonomy of the robot. In this study, we aim at applying deep learning techniques to allow our humanoid robot ARMAR-IIIa to explore the semantics of the scene and actions by processing the perceived raw RGB-D image streams. The proposed work will yield localization and recognition of possible functional areas from an unseen environment.

In this project, the candidate has to conduct various experiments on already existing publicly available datasets. First, a multi-layer Convolutional Neural Network (CNN) has to be implemented for scene functionality anlysis. To train the network, different existing labeled datasets will be considered. The final performance of the system has to be tested with our humanoid robot ARMAR-IIIa in a kitchen environment. Applicants must be interested in deep learning and should possess good programming skills either in Python, C/C++, or Matlab.

Figure 1: Deep learning for semantic scene and action perception. On the left the humanoid robot ARMAR-IIIa observes a scene including a chair, walls and cabinets. The aim here is to extract and label individual scene segments. For instance, the brown segment (see image on the right) represents the chair shown in the middle.