Online Recognition of Daily-Life Movements

D. Gehrig, A. Fischer, H. Kühne, T. Stein, A. Wörner
H. Schwameder and T. Schultz
Universität Karlsruhe (TH), Germany

The task of coaching a robot includes a variety of processes. These processes range from active sensor control to abstract applications like the recognition of the current situation. One of these processes is the recognition of human movements. For the robot it is essential to know the motions of humans who are coaching it or who are part of the scenario that should be analysed by the robot. Our goal is to develop such a motion recognition system for the humanoid robot ARMAR as part of the CRC 588 "Humanoid Robots - Learning and Cooperating Multimodal Robots". In our current research we use joint angle trajectories of human motions for the recognition of the motion sequences based on Hidden Markov models (HMMs). The joint angles are caculated using marker-based tracking of the observed human.

In our current work we focus on human motions as they appear in a kitchen and food preparation scenario. We investigate 8 different motion sequences consisting of motion units like picking and placing objects or using them, e.g. cutting an apple. For the data acquisition a human subject was asked to perform these motion sequences while standing at a fixed position facing a table with the objects lying on the table at fixed positions. At the beginning and between some motion units the subject is going into a neutral position with both hands resting on the table.

For the tracking of the human motions light reflecting markers were attached to the subjects body. 10 Vicon infra-red cameras were used to track the marker positions in space. For the recognition process, the marker positions are directly streamed into our software, where the corresponding joint angles are calculated. For the angle calculation we use a rigid upper body multibody model of the human skeleton, which is transformed to best fit the marker positions for each time step. For the training of the recognizer we recorded 400 motion sequences of one subject.

We use 3-state continuous left-to-right HMMs to model each motion unit. Each human motion sequence is modeled as a concatenation of these motion units. For the recognition of the motions we use the one pass IBIS decoder, which is part of the Janus Recognition Toolkit. Decoding was carried out as a time-synchronous beam search guided by a context free grammar. For the initialisation of the system about two third of the training data has been manually segmented into the mentioned motion units.

The data acquisition, preprocessing and recognition have been optimized for online recognition. We are able to recognize human motions online, while they are performed by the subject. In particular the data acquisition, the joint angle reconstruction and the motion recognition itself are incorporated into a single framework and we are able to recognize motions with a frame rate of 20 Hz on a standard PC. On a test set of 80 motion sequences, which have not been seen during the training process, we achieved a sequence recognition accuracy of 100 %.

For the future we are interested in dealing with more variance in the performed motions. On the one hand we will investigate the variability in persons movements. We plan to investigate inter- and intraindividuell differences in human motions especially according to velocity and precision, which partly depend on the temporal context of the motion units. The results are a basis for the improvement of the recognizer.

On the other hand we will enhance our motion recognition system. We will investigate how to build a recognizer that is able to recognize movements of more than one person. We also plan to make our motions more natural by a higher number of possible object positions and by the avoidance of going into neutral position between motion units. The neutral positions were quite helpful for modelling the motion units independent of their context in time and helped for the segmentation of the training data, but did not represent natural behaviour.

Recording the training and test data with Vicon is fast and reliable, which enables us a fast development of the motion recognizer. In the near future, we plan to test and extend our models and algorithms on joint angle data that is going to be captured based on video streams. For the data acquisition the same stereo camera system will be used, that is part of our robot ARMAR.