Interacting with robots

Frank Wallhoff, Christoph Mayer, and Tobias Rehrl

Multimodal input channels show great promise for cognitive systems.

Everyday human-human interactions rely on a large number of different communication mechanisms, including spoken language, facial expressions, body pose, and gestures, thus allowing humans to pass large amounts of information in a short time. In contrast, traditional human-machine communication via mouse, keyboard, and screen is often nonintuitive and requires time-consuming operations, even for trained personnel (see Figure 1). Because of this inconvenience, improving human-machine interactions in general (and human-robot interactions in particular) is currently of great interest.

Several research facilities are dedicated to realizing natural human-robot interactions by equipping technical systems with a degree of cognition that allows for more intelligent and useful robotic reactions in response to the surroundings, in particular to a human interaction partner. The main idea that forms the basis of research in the European Commission’s excellence cluster ‘Cognition for Technical Systems’ (CoTeSys) is to apply insights gained from research in biology, psychology, and neuroscience to robotics, aimed at creating cognitive systems that are capable of adapting to their environment rather than following predefined rules. We approach human-machine interactions based on two scenarios characterized by very different challenges, ambient living and advanced robotics in an industrial context. CoTeSys follows an interdisciplinary approach, with researchers from diverse areas (including computational linguistics, computer science, electrical engineering, and psychology) jointly tackling the challenge of cognitive systems.

Humans do not only rely on a single communication channel such as spoken language. Therefore, cognitive systems must be able to comprehend humans in a multimodal manner. Although spoken language is used to pass on large amounts of information, nonverbal interactions, such as gestures or facial expressions, contain valuable context information and contribute significantly to everyday human-human communications. They demonstrate a human’s emotional state, show agreement or disagreement, provide various greeting signs, and may augment or even replace information passed on by spoken language. Therefore, our research focuses on these communication signals.

We set up a real-time-capable gesture-interaction interface for

Continued on next page
human-robot interaction to evaluate gestures (head, hand) as well as facial expressions (see Figure 2).

To enable multimodal data processing, a suitable communication framework is generally required, because data from different input sources must be processed in both real-time-capable and temporally synchronized manners. Therefore, we integrated the Real-Time Database (RTDB)—a high-performance communication platform developed at the Technical University of Munich—in the context of cognitive automobiles. The RTDB is a sensory buffer that records varying data during a given time period and makes this available to different modules (shared-memory characteristic). In addition to low computational-processing overheads, different modules can process the same data without any blocking effects (see Figure 3). The interfaces between the writer module (making data available in the RTDB) and the reader module (extracting data) are strictly defined, which enables exchanging parts of the system without the need for adjustment of other system components.

Our system takes a generic approach. We implement a recently published object-detection approach to detect human faces and adapt a skin-color model with regard to the face image obtained. To constrain hand-gesture recognition, we define a region of gesture action for feature extraction. To obtain information about face shape and head pose, we integrate the publicly available Candide-III face model. Temporal changes in the model parameters indicate facial expressions or head gestures. We apply continuous hidden-Markov models to determine dynamic hand and head gestures, while support-vector machines are trained to recognize facial expressions. To date, the real-time capable framework can support human-robot interaction in two scenarios, an assistive household and an industrial hybrid assembly station. In the assistive-household scenario, humans are an integrated part of the environment and this setup is, therefore, a good starting point for our research. Our objective for integration of gesture recognition in the industrial context is mainly driven by the fact that reliable speech recognition is not always available, since ambient and unpredictable noise within a factory hinders the speech-recognition process.

We have presented our demonstrator at trade fairs, scientific conferences, laboratory tours, and on TV. Despite the great potential of the approach handling multimodal input data, there is still room for improvement (which is the focus of our continued research). To date, we have been treating the outputs of the different modalities separately. However, the data can be fused early or late. The major drawback of our approach is that classifiers are learned offline in advance. However, online learning of gestures and facial expressions would enable adaptation to either a single human or a specific interaction environment.

Figure 3. Typical situation in an industrial human-robot interaction scenario. The artificial co-worker must perceive all relevant actions in the environment in a multimodal manner. From this input, the robot controller must decide (in real time) the next action that does not violate any given security constraint.

Author Information

Frank Wallhoff  
Technical University of Munich  
Munich, Germany

Frank Wallhoff is head of the Interactive Systems Group. His research interests cover cognitive and assistive systems. In addition, he is principal investigator of the Cluster of Excellence ‘Cognition for Technical Systems’ and coordinator of a number of international projects focusing on human/machine interaction and assistive technologies.

Christoph Mayer  
Intelligent Autonomous Systems Group  
Technical University of Munich  
Munich, Germany

Christoph Mayer received his Dipl.-Inf. degree in computer science from the Technical University of Munich in 2007 where he started at the image understanding and knowledge based systems group in January 2008. In January 2010, he joined the Intelligent Autonomous Systems Group. His research interests relate to image understanding, where he investigates face-model fitting and facial-expression recognition.

Continued on next page
Tobias Rehrl
Institute for Human-Machine Communication
Technical University of Munich
Munich, Germany

Tobias Rehrl received his Dipl.-Ing. degree in electrical engineering from the Technical University of Munich in 2007. Since March 2008 he has been working as a researcher affiliated with the Institute for Human-Machine Communication. His research interests focus on image processing, in particular tracking, as well as applying graphical models in the field of pattern recognition.

References