Sensor-based Situated, Individualized, and Personalized Interaction in Smart Environments

S. Hämmerle, M. Wimmer, B. Radig, M. Beetz

Informatik IX
Technische Universität München
Boltzmannstr. 3
85748 Garching bei München
[haemmerl, wimmerm, radig, beetz]@in.tum.de

Abstract: Smart environments are sensor equipped areas that know about their environment thus being able to adapt to the user. We present *sHOME*, a multi-agent based platform for integrating situated, individualized, and personalized information. *sHOME* acquires sensor data to determine the user's identity, his location, his gesture, and natural language commands and stores it in a central knowledge base.

1 Motivation

Within smart environments computer systems are equipped with various input devices (cameras, microphones …) in order to acquire knowledge about their environment. This information describes the current situation including the number and the identities of the persons present, spoken sentences, executed gestures, and even their mood. Smart environments need to manipulate the environment by actuators such as increasing the volume of the TV. For communication with the user the smart environment uses written text and spoken language. Ideally people interact with smart environments just like they do with humans using natural language, gestures, and mimics.

Smart environments require situated, individualized, and personalized information to give optimal support to the user. Two kinds of information can be stored within the computer system in advance: information about the user himself like his age or gender (personalization), as well as the user's preferences, like his favorite TV channel (individualization). Information about the current situation has to be acquired continuously by sensors.

This document reports the use of several state-of-the-art techniques that are able to gather information about the current situation. We designed a platform named *sHOME* (Smart Home, Office, and Meeting Environment) that supports the acquisition and combination of this information. Any information is translated into semantic content. We demonstrate some scenarios, showing its range of application.

2 Related work

Many research groups try to explain various kinds of sensor data in order to automatically analyze the current situation. Kwolek [Kw03] uses a stereo vision camera mounted on a robot to detect the user's position and react to his commands that are given via arm posture. Francois [Fra04] detects human posture and activity using a camera mounted within a room. Stiefelhagen et al [Stie99] record a meeting scenario and detect the talking person by analyzing the focus of attention of the other people combined with audio information.

Recently, a few research projects combine different techniques for situation detection and integrate them within a smart environment. There are two projects that integrate a smart environment into a classroom. The Intelligent Classroom [FH01] detects situations like going to the board and write something down. It adaptively supports the lecturer by controlling the slides via voice or gestures. The project eClass [BA98] also serves as audio-visual assistant but focuses on supporting the pupils providing them with information about the current topic on their PDA.

The Context Broker Architecture (CoBrA) is an agent-based architecture for supporting context-aware systems. The context broker is responsible for maintaining and sharing situated, individualized, and personalized information for a community of agents, services, and devices and provides privacy protection by enforcing the policy rules. A semantic representation is used for the upcoming data [Ch04b].

3 Our Setup: *sHOME*

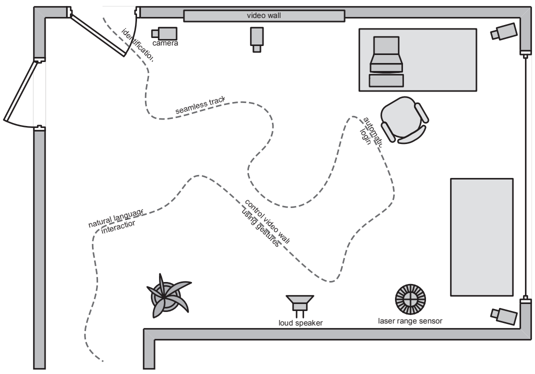

Figure 1: A demo scenario.

With our project *sHOME* we aim at detecting the situated context within a room and making use of that information. We equipped a room at our office with cameras, laser range sensors, headset microphones, loud speakers, a beamer for a video wall, and a computer workstation (see). We can identify persons within that room thus we can make use of their individual preferences and personalized settings which we manually specify in advance.

In this example scenario we show the benefit of personalization, individualization, and situation detection. A camera observes the entrance area and visually identifies persons that enter the room. To adapt to the person the system immediately loads personalized and individualized information such as his preferred volume settings for the loud speaker. *sHOME* welcomes the person in a personalized matter (“Hello/Good morning/…”). Furthermore individual privileges and restrictions are set. As long as he stays inside of the room his position is detected by a laser range sensor. When sitting down at the computer workstation, the person is provided with his personal desktop without having to log in manually. Using the integrated multimedia functionality, the user can listen to music and watch movies that can be controlled via gestures and natural language. (see *sHOME*-Video [JW05])

3.1 The software architecture

Complex calculations to interpret sensor data such as gesture recognition are done by separate modules (see 3.2) that forward the results to an associated agent. Sourcing out that functionality into external modules allows for integrating currently existing programs and the use of various programming languages. Because of its inherent distribution we built *sHOME* upon a multi-agent framework. We focused on exchanging semantic information (knowledge) between the agents rather than raw data. This guarantees consistency within the high number of involved components. As you can see in we created a specific agent for each corresponding module. The agent’s job is to translate the sensor data into semantic information and forward it.

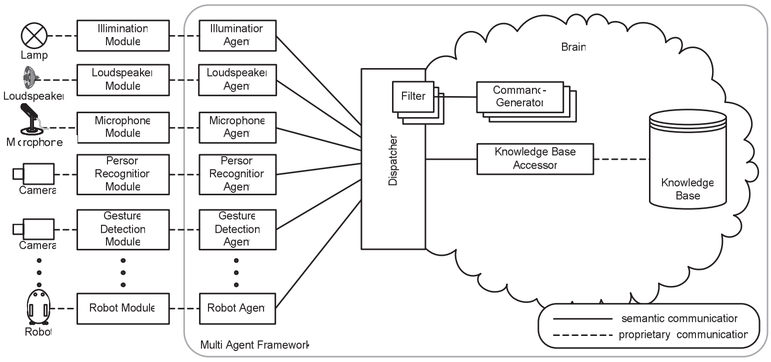

Figure 2: The components of the *sHOME*.

All acquired information is sent to a central entity, the so-called brain. The brain is responsible for gathering and storing the information, making decisions, and reacting to events. It contains a knowledge base which has been manually filled with information about the employees, projects, and teachings at our chair. It also includes preferences and roles which are necessary for individualization and personalization. Extended with rules new, previously not existing, knowledge can be inferred. E.g. A is boss of B and B is boss of C thus A is boss of C

Both the knowledge base and the content of the exchanged messages have been built up on the Standard Ontology for Ubiquitous and Pervasive Applications (SOUPA) [Ch04a]. SOUPA has been written in the Web Ontology Language (OWL) which is an emerging standard for knowledge representation. For querying the knowledge base and reasoning about its content we use Stanford's OWL query language (OWL-QL). So far our knowledge base does not handle situated preferences and use preference repositories for personalization as proposed by [HK04]. We are currently investigating how our knowledge base should be extended to deal with these aspects.

The dispatcher promotes each incoming message into the knowledge base. If the message matches with a filter it is forwarded to the corresponding command generator. The command generator agents react on events and make decisions. There are command generators for various tasks such as responding on natural language input.

3.2 Handling the sensor input

The persons' identities can be detected using various techniques such as RFID reader, finger print readers, or iris scan. Currently *sHOME* uses face recognition because this is convenient and non-intrusive. Faces in the camera images are identified by a classifier which has been trained with the members of our chair in advance.

The persons' locations are detected by a laser range sensor that was horizontally installed 30 cm above the floor. At this height legs can be detected very well, without any concern if the person is sitting, standing, or walking. But it cannot detect a person or an object that is located behind another object. Once a person's identity has been recognized the person's position is tracked seamlessly. That makes it possible to permanently know where a person is.

Gestures can be used to communicate with *sHOME*. We defined a set of gestures that can be executed and remembered easily. In order to allow an intuitive and ambient handling the meaning of the gestures can be set by the user (individualization). Simple yes/no questions can be answered by the head gestures nodding and shaking.

Focus of attention: Persons can pay attention to their conversation partners or look at devices within the room. Depending on the gaze of the person natural language commands like "switch off that light" demands for different behavior of *sHOME*. If a person is currently talking to somebody else his words are not intended to be a voice command at all. *sHOME* detects the person's focus of attention using camera images. A face model is adjusted in order to fit to the face in the image and describes the orientation of the head.

Human emotion is primarily expressed through mimics. Mimics can be observed as small muscle motion within the face. We created a deformable face model that can adapt to those movements. Observing those movements over a short time human mimic can be inferred. Right now we are able to detect neutral faces, laughing, and surprise. [FD03]

Natural language is the predominantly used channel of human to human interaction and carries the most information. We integrated the CMU Sphinx IV natural language recognition engine which requires relatively low computational power. For the use in sHOME we defined our own set of commands.

4 Conclusion and Future Work

In this paper we have presented *sHOME*, a sensor equipped room which we use for demonstrating personalization, individualization, and situation detection. The advantage of *sHOME* is that information about the current situation is acquired automatically. Due to the semantic representation of the information complex queries can be answered. Because of the modular design new sources of situated information can be attached.

At the moment, personalization information such as the user preferences is specified manually. In the future, this information could automatically be created, modified, and refined during the interaction with the user. Furthermore, we will deploy our system into some more rooms. Like that we can demonstrate more complex scenarios such as tracking people across room borders. Currently the main communication channels within smart environments are natural language and dedicated gestures. In the future this can be completed by integrating the recognition of body language.

5 Literature

[BA98] Jason A. Brotherton and Gregory D. Abowd, Rooms take note: Room takes notes, AAAI 1998 Spring Symposium on Intelligent Environments. AAAI TRSS-98-02, 1998

[Ch04a] H. Chen et al., SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications, International Conference on Mobile and Ubiquitous Systems, August 2004.

[Ch04b] H. Chen et al., Intelligent Agents Meet the Semantic Web in Smart Spaces, IEEE Internet Computing, November 2004.

[FD03] S. Fischer, S. Döring, M. Wimmer, A. Krummheuer, Experiences with an Emotional Sales Agent, Workshop on Affective Dialogue Systems. pp 309-312, Kloster Irsee, Germany, Juni 2004.

[FH01] David Franklin and Kristian Hammond, The intelligent classroom: providing competent assistance, Fifth International Conference on Autonomous Agents, pp 161-168, 2001

[Fra04] B. Boulay et. al., Human Posture Recognition in Video Sequence, Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp 23-29, Oct 2003, Nice

[HA93] D. Huang, F.Alleva, H.W. Hon, M. Y. Hwang, K. F. Lee, and R. Rosenfeld, The Sphinx-II Speech Recognition System: An Overview, Computer, Speech and Language, 1993

[HK04] S. Holland, W. Kießling, Situated Preferences and Preference Repositories for Personalized Database Applications, 23rd International Conference on Conceptual Modeling (ER 2004). S. 511-523, Shanghai, China, Nov. 2004.

[JW05] A. Jung, M. Wimmer: http://wwwradig.in.tum.de/people/wimmerm/sHOME-Video

[Kw03] Bogdan Kwolek, Visual System for Tracking and Interpreting Selected Human Actions, Journal of WSCG, Vol.11, No.1., WSCG'2003, February, 2003, Plzen, Czech Republic.

[SW94] B. Schiele, A. Waibel, Gaze Tracking Based on Face-Color. Format-Verlag, Bonn, 1999; S. 135-162

[Stie99] R. Stiefelhagen, J. Yang, A. Waibel, Modeling Focus of Attention for Meeting Indexing, Proceedings of ACM Multimedia '99, pp 3-10, Oct 1999, Orlando, Florida, USA