Towards On- and Off-line Search, Browse and Replay of Home Activities

Anton Nijholt
Human Media Interaction, University of Twente
PO Box 217, 75000 AE Enschede
The Netherlands
anijholt@cs.utwente.nl,
WWW home page: http://hmi.ewi.utwente.nl/~anijholt

Abstract. Ambient Intelligence research is about ubiquitous computing and about social and intelligent characteristics of computer-supported environments. These characteristics aim at providing inhabitants or visitors of these environments with support in their activities. Activities include interactions between inhabitants and between inhabitants and (semi-) autonomous agents, including mobile robots, virtual humans and other smart objects in the environment. To provide real-time support requires understanding of behavior and activities. Clearly, being able to provide real-time support also allows us to provide off-line support, that is, intelligent off-line retrieval, summarizing, browsing and even replay, possibly in a transformed way, of stored information. Real-time remote access to these computer-supported environments also allows participation in activities and such participation as well can profit from the real-time capturing and interpretation of behavior and activities performed supported by ambient intelligence technology. In this paper we illustrate and support these observations by looking at results obtained in several European and US projects on meeting technology. In particular we look at the Augmented Multi-party Interaction (AMI) project in which we are involved.

1 Introduction

Environments equipped with Ambient Intelligence technology provide social and intelligent support to its inhabitants. The majority of ambient intelligence research is on providing support to individuals living or working in these smart environments. However, in home and office environments we have also people interacting with each other and interacting with smart objects (e.g., a mobile robot, furniture, intelligent devices, and virtual humans on ambient displays). Cameras, microphones and other sensors can be used to detect and capture such activities. Can the

environment, using this sensorial input, support this multi-party interaction, e.g. in a home environment, as well?

Looking at smart environments from the point of view of supporting multi-party interaction adds some interesting research issues to the area of ambient intelligence research. Firstly, in order to be able to provide support, the environment is asked to understand the interactions between its inhabitants and between inhabitants and the environment or smart and maybe mobile objects available in the environment. Although we see the development of theories of interaction and behavior, these theories are rather poor from a computational point of view and therefore they hardly contribute to the design of tools and environments that support activities of human inhabitants. Hence, the need for computational theories of behavior and interactions needs to be emphasized. A second research issue that needs to be mentioned is the real-time monitoring of activities, the on-line access to information about activities taking place and also the on-line remote participation in activities or influencing activities in smart environments. The third research issue concerns the off-line access to stored information about activities in smart environments. This latter issue may involve retrieval, summarization, replay and browsing.

Certainly, not all three research issues need to be considered for every type of smart environment. Sometimes we are only interested in providing real-time support to an individual entering an ambient intelligence environment. Sometimes we just want to monitor what is happening and having an alert when something unusual occurs. Sometimes we want to know what activities were there when we were not present. Sometimes we need to retrieve, browse or replay previously stored information about activities in the past in order to support current activities.

There is one important domain of application of ambient intelligence technology where all these research issues play an important role. This is the domain of meetings supported by smart environment technology. In this domain it is useful to provide support during the meeting, it is useful to allow people who can not be present to view what is going on, it is useful to allow people to remotely participate and it is useful to provide access to captured multimedia information about a previous meeting, both for people who were present and want to recall part of a meeting and for people who could not attend.

The aim of this paper is to look at the way results from research and development done in the context of some large research projects on the design and development of meeting support technology (smart meeting rooms, remote meeting participation, distributed meetings, distributed collaborative work spaces, etc.) can be explained and explored in the context of smart home environments.

In section 2 of this paper we look at ambient intelligence in home environments and extend existing views in order to include multi-party interaction support and replay of events. In section 3 of this paper we discuss the research issues in several projects dealing with the development of meeting support technology. We explain and review the research approaches from a point of view that allows exportation to other research and application areas. In section 4 we extend these views and approaches to (remote) meeting support such that it becomes clear that topics such as visualization, virtual reality and embodied agents (virtual humans).can play important roles in providing not only meeting support, but also, with appropriately equipped smart home environments, to support (1) multi-party interaction and joint

activities of family members (including virtual pets and virtual humans), (2) real-time monitoring and participation in such activities, and (3) retrieving, browsing, and replaying of previously captured and stored information about activities that took place in a particular environment. Section 5 contains conclusions and has observations about future research.

2 Social and Intelligent Home Environments

Whatever kind of situation we are in, when 'ambient intelligence' in one or other way is able to support our activities we can be happy with it. Maybe the activities can be done more efficiently due to this support or they can become more enjoyable. Do we want to look back at activities, do we want to retrieve information about previous activities or do we want to experience these activities again, maybe from an other view point or being in an other's person skin?

Our viewpoint is that there are lots of reasons to want to look back on a previous activity in which we or our friends and relatives were involved. This is certainly obvious when looking at a meeting event. We always do, trying to remember what happened, what was said and what decisions were taken for what reasons. Traditionally there are minutes of a meeting, participants have their own notes and there is other material that can be consulted (agenda, list of participants, documents, presentations). More and more we see audio and video recordings of meetings appear in order to be able to back to a certain moment during a meeting. This makes clear that meetings differ from spontaneous gatherings, from family gatherings and, generally, meetings and joint activities between friends, relatives and family members. Meetings are structured and certain goals are defined in advance.

Hence, a meeting differs from joint activities in a home environment, but also in home environments meeting support technology that is now developed in some large European projects can play useful roles. The home environment can ask for real-time support for activities that take place, sometimes it can be useful or enjoyable to remotely take part in home activities and sometimes we would like to experience in some or other way an important moment again. Presently this is done with diaries, photo albums and video collections. Web providers make it already possible to share these collections with others. Personal archives are made accessible for others and personal notes and thoughts appear in blogs on the web. This can be considered as a first step to a continuous registration of events in social environments [1] and at the same time to technology that makes it possible to search, browse and replay such information or allow to get immersed in this information (see also [2]).

Currently, most ambient intelligence technology that is being developed concerns applications as home environment control and automation. Personal entertainment, health care and security are other application areas. In our view we should also look at events that involve multi-party interaction for which real-time support is useful and where support requires some high-level interpretation (in contrast with turning on the lights when someone enters the room). This interpretation allows also for off-line intelligent search in the stored information, the development of intelligent browsing tools and multimedia presentation of the information. Among the possibilities for multimedia presentation we include ways of replaying, probably in a

transformed and manipulated way of home activities (family meetings, visits of relatives, playing with children, a birthday party, a wedding, just an evening at home with everyone doing usual things, preparing a dinner in the kitchen, et cetera). Being able to interpret, search, browse and replay recorded meeting data is part of the European AMI (Augmented Multi-party Interaction) project. Having a (mixed reality) 'album' of important events is one of the streams (*My Life Album*) of the *IntoMyWorld* candidate Presence II project [3]. Among the examples that are mentioned is the possibility to allow people to re-immerse themselves in their own weddings. In this paper an attempt is presented to bring these approaches together.

3 The AMI (Augmented Multi-party Interaction) Project

3.1 General Background and Introduction

By looking at the earlier mentioned AMI project we want to make clear that technology obtained in multi-party interaction research as is now becoming available, can be usefully employed in the context of other smart environments. The AMI[1] project builds on the earlier M4 project (Multi-Modal Meeting Manager). Both projects are concerned with the design of a demonstration system that enables structuring, browsing and querying of archives of automatically analyzed meetings. The meetings take place in a room equipped with multimodal sensors. Multimedia information captured from microphones and cameras are translated into annotated multimedia meeting minutes that allow for retrieval, summarization and browsing. The result of the M4 project was an off-line meeting browser.

More than in M4, in the recently started AMI project attention is on multimodal events. Apart from the verbal and nonverbal interaction between participants, many events take place that are relevant for the interaction and that therefore have impact on their communication content and form. For example, someone enters the room, someone distributes a paper, a person opens or closes the meeting, ends a discussion or asks for a vote, a participants asks or is invited to present ideas on the whiteboard, a data projector presentation is given with the help of laser pointing and later discussed, someone has to leave early and the order of the agenda is changed, etc. Participants make references in their utterances to what is happening, to presentations that have been shown, to behavior of other participants, etc. They look at each other, to the person they address, to the others, to the chairman, to their notes and to the presentation on the screen, etc. Participants have facial expressions, gestures and body posture that support, emphasize or contradict their opinion, etc.

To study and collect multimodal data smart meeting rooms are maintained by the different research partners. They are equipped with cameras, circular microphone arrays and, recently introduced, capture of whiteboard pen writing and drawing and note taking by participants on 'electronic paper'. Participants also have lapel microphones and cameras in front of them to capture facial expressions.

[1] AMI started on 1 January 2004 and has duration of three years. It is supported by the EU 6th FP IST Programme (IST IP project FP6-506811).

3.2 AMI: From Signal Processing to Interpretation

The meeting support application researched in the AMI project [4] requires the development of tools that take into account the meeting context. Rather than zooming in on constraining general methods of detecting and interpreting events in physical environments, we have a bottom-up approach starting with observed events in meeting environments and attempting to model and explain them using more general observations on theories of verbal and nonverbal communication.

Models are needed for the integration of the multimodal streams in order to be able to interpret events and interactions. These models include statistical models to integrate asynchronous multiple streams and semantic representation formalisms that allow reasoning and cross-modal reference resolution. Apart from the recognition of joint behavior, i.e., the recognition of group actions during a meeting, there is also the recognition of the actions of individuals, and the information fusion at a higher level for further recognition and interpretation of the interactions.

When looking at the actions of the individuals during a meeting several useful pieces of information can be collected. First of all, there can be person identification using face recognition. Current speaker recognition using multimodal information (e.g., speech and gestures) and speaker tracking (e.g., while the speaker rises from his chair and walks to the whiteboard) are similar issues. Other, more detailed but nevertheless relevant meeting acts can be distinguished: for example, recognition of individual meeting actions by video sequence processing.

Presently models, annotation tools and mark-up languages are being developed in the project. They allow the description of the relevant issues during a meeting, including temporal aspects and including low-level fusion of media streams. In our part of the project we are interested in high-level fusion, where semantic/pragmatic (tuned to particular applications) knowledge is taken into account (see e.g. [5]). I.e., we try to explore different aspects of the interpretation point of view. We hope to integrate recent research in the area of traditional multimodal dialogue modeling. These issues will become more and more important since models, methods and tools that need to be developed in order to make this possible can be used for other events taken place in smart and ambient intelligence environments as well.

4 Towards Virtual Reality Representations and Replay

In our research we have looked at capturing meeting activities from an image processing point of view and at capturing meeting activities from a higher-level point of view, that is, a point of view that allows, among others, observations about dominance, focus of attention, addressee identification, and emotion display. We studied posture and gesture activity, using our vision software package. A flock-of-birds package was used to track head orientation of some of our 4-party meetings. It allowed us to display animated representations of meeting participants in a (3D) virtual reality environment [6]. In this environment visualized events can be augmented with meta-observations provided by support agents and displayed in the virtual environment. This is illustrated in Fig. 1.

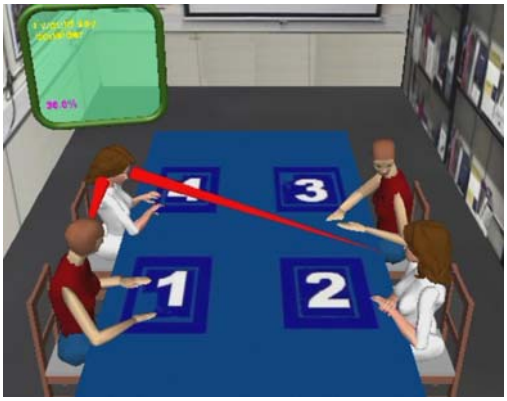

Fig. 1. The virtual meeting room showing gestures, head movements, speech transcript, addressee(s) and the percentage a person has spoken until that moment

Even more attractive is it to have meetings represented in a virtual meeting room (VMR), where participants do not all share the same physical space. We introduced a prototype version of a distributed meeting room set-up. This set-up [7] allows the connection of several inhabited smart meeting rooms and the representation of the participants and their activities in a shared virtual environment, made accessible for participants (and observers) in real-time. It allows the participants to take part in the meeting, perceiving the verbal and nonverbal communication by other participants through their avatars, from their assigned position around the meeting table. As shown in Fig. 2, also in this distributed version we can add meta-information about the meeting and its progress to the visualization of the virtual room.

The technology used within the DVMR experiment differs substantially from normal video conferencing technology. Rather than sending video data as such, this data is transformed in a format that enables analysis and transformation. For the DVMR experiment the focus was on representing poses and gestures, rather than, for example, facial expressions. Poses of the human body are easily represented in the form of skeleton poses [8], essentially in the same format as being used for applications in the field of virtual reality and computer games. Such skeleton poses are also more appropriate as input data for classification algorithms for gestures.

Another advantage for remote meetings, especially when relying on small handheld devices, using wireless connections, is that communicating skeleton data requires substantially less bandwidth than video data. A more abstract representation of human body data is also vital for combining different input channels, possibly using different input modalities. Here we rely on two different input modalities: one for body posture estimation based upon a video camera, and a second input channel using a head tracker device. Although the image recognition data for body postures also makes some estimation of the head position, it turned out that using a separate head tracker was much more reliable in this case.

The general conclusion is, not so much that everyone should use a head tracker device, but rather that the setup as a whole should be capable of fusing a wide variety of input modalities. This will allow one to adapt to a lot of different and often difficult situations. In the long run, we expect to see two types of environment for

remote meetings: specialized meeting rooms, fully equipped with whatever hardware is needed and available for meetings on the one hand side, and far more basic single user environments based upon equipment that happens to be available. The capability to exploit whatever equipment is available might be an important factor for the acceptance of the technology. In this respect, we expect a lot from improved speech recognition and especially from natural language analysis. The current version of the virtual meeting room requires manual control, using classical input devices like keyboard or mouse, in order to look around, interact with objects etcetera. It seems unlikely that in a more realistic setting people that are participating in a real meeting would like to do that. Simpler interaction, based upon gaze detection but also on speech recognition should replace this situation

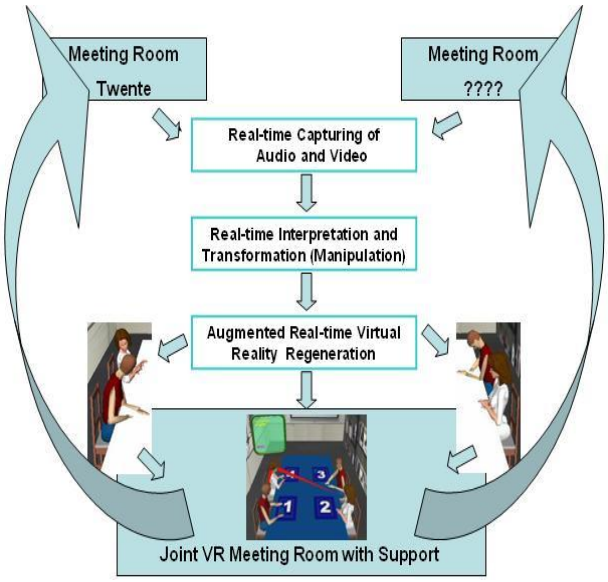

Fig. 2. Capturing, manipulation and re-generation of activities in remote locations in a joint virtual meeting room

5 Conclusions

Home automation is important, but providing real-time support to inhabitants during their activities is important as well. This real-time support requires interpretation of home activities. In many of these activities we have to deal with multi-party interaction. That is, there are verbal and nonverbal interactions between the human inhabitants of the environment. Moreover, with the introduction of mobile robots, smart objects and virtual embodied agents displayed on walls and objects, the multi-party members will also include these artificial and pro-active agents. The environment needs some understanding of such interactions and therefore we need to look for models for multi-party verbal and nonverbal interaction.

Meetings are rather controlled events and therefore they are a more acceptable target for preliminary research in this direction. We looked at the approaches and preliminary results obtained in the European AMI project on smart meeting environments. In this project real-time support is only one of the objectives. Rather the emphasis is on querying and browsing the multimedia information that is captured using various types of sensors. Being able to replay in one or other form of a meeting is an interesting objective. These additions to real-time support are useful in home environments as well. Apart from real-time support to home inhabitants and real-time remote access from other smart environments, it also allows intelligent querying, browsing and replay of previous interesting events. From detecting rather straightforward events as entering a room, being in the proximity of a certain object or identifying a person in the room, to the interpretation of events in which more persons are involved is a rather big step. However, in AMI and other large EU projects we now see, as discussed here, that small steps in this direction are taken.

Acknowledgements

I want to thank Job Zwiers, Rutger Rienks, Hendri Hondorp and Ronald Poppe for their research contributions. Jan Peciva from the Technical University of Brno helped with the realization of the distributed virtual meeting room. This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-147).

References

1. M. Deutscher, P. Jeffrey & N. Siu. Information capture devices for social environments. In: Proceedings *EUSAI 2004*, LNCS 3295, Springer-Verlag Berlin Heidelberg, 2004, 267-270.
2. S. Vemuri & W. Bender. Next-generation personal memory aids. *BT Technology Journal*, Vol. 22, No 4, October 2004, 125-138.
3. P. Turner, S. Turner & D. Tzovaras. Reliving VE day with schemata activation. In: Proceedings *8th International Workshop on Presence: Presence 2005*. M. Slater (ed.), London, 2005, 33-38.
4. I. McCowan, D. Gatica-Perez, S. Bengio, D. Moore, H. Bourlard. Towards Computer Understanding of Human Interactions. In: *Ambient Intelligence*, E. Aarts et al. (Eds.), LNCS, Springer-Verlag Heidelberg, 235 - 251.
5. A. Nijholt. Multimodality and Ambient Intelligence. In: *Algorithms in Ambient Intelligence*. W.F.J. Verhaegh, E.H.L. Aarts & J. Korst (eds.), Kluwer, Boston, 2003.
6. A. Nijholt, J. Zwiers & J. Peciva. The Distributed Virtual Meeting Room Exercise. In: Proceedings ICMI 2005 Workshop on Multimodal multiparty meeting processing, A. Vinciarelli & J-M. Odobez (eds.), Trento, Italy, October 2005, 93-99.
7. A. Nijholt. Meetings in the Virtuality Continuum: Send Your Avatar. In: Proceedings 2005 International Conference on CYBERWORLDS, T.L. Kunii, S.H. Soon & A. Sourin (eds.), IEEE Computer Society Press, Los Alamitos, USA, November 2005, Singapore, 75-82.
8. R. Poppe, D. Heylen, A. Nijholt, & M. Poel. Towards real-time body pose estimation for presenters in meeting environments. *Proc. 13th Intern. Conference in Central Europe on Computer Graphics, Visualization and Computer Vision.* V. Skala (Ed.), Plzen, Czech Republic, 2005, 41-44.