Recognition and Reasoning in an Awareness Support System for Generation of Storyboard-like Views of Recent Activity

Datong Chen & Hans-Werner Gellersen
Telecooperation Office (TecO), University of Karlsruhe
{charles, hwg}@teco.uni-karlsruhe.de

ABSTRACT
Awareness support system are based on formal and specific context information such as location, or on video-mediated general context information such as a view into a remote office. We propose a new approach based on fusion of these different kinds of context information. In this approach we distinguish white box context, used by the awareness system for reasoning, and black box context, which can only be interpreted by humans. Our approach uses a variety of perception techniques to obtain white box context from audio and video streams. White box context is then used for further processing of context information, for instance to derive additional context. It is further used to generate a storyboard-like multimedia representation of collected and extracted context information. This storyboard provides a condensed view of recent activity to collaboration partners.

Keywords
Collaboration awareness, context-awareness, awareness support systems, context recognition, groupware

INTRODUCTION

Awareness support systems are designed to provide distributed people with context information for collaborative activity. The idea is to make up for the lack of cues that people use in face-to-face settings to stay aware of their colleague's availability for interaction and collaboration. Cues such as whether a colleague appears to be very busy guide the social coordination of collaborative work. People may use individual cues spontaneously, for instance to join a conversation they overhear, but often exposure to cues over a longer period is required to assess a particular situation, for instance to decide who in a group best to approach for help with a collaborative task.

In distributed collaborative work, people have only limited context information available to guide collaboration, and to decide for instance when to initiate a particular collaborative task. Computer-based tools used to support distributed collaborative work, for instance shared applications and shared workspaces, provide only formal awareness of other people's activities. Formal awareness is restricted to activity mediated within the shared application and neglects any context information beyond, such as whether a colleague also has to attend other tasks. Distributed real time communication can be used to explicitly communicate additional context, and to add social coordination to formal collaboration. The problem remains that people in different locations lack the cues based on which they can decide about initiation of such communication [3]. This problem is addressed by awareness support systems, designed to provide collaborating people with awareness of each other and with cues to guide social coordination of collaborative work.

A variety of awareness support systems has been discussed in the community [3,10,12,13]. In addition to formal awareness, these systems promote awareness of informal context. Generally, the approach is to continuously capture and transmit context information in real time. Most systems capture and transmit context information that is not interpreted but just mediated by the system, for example audio and video streams. For such context information we use the term *black box context.* Some awareness systems interpret collected context information to obtain and transmit *white box context.* For example, active badge systems transmit location information as white box context, suited for further processing in a groupware system.

In this paper we propose an awareness support system that is based on the following concepts:

- Fusion of context information obtained from different sources, in particular fusion of information obtained from specialized sensors with more general information obtained with audio/video techniques.
- Explicit notion of black box context vs. white box context.
- Use of perception methods to obtain white box context as foundation for reasoning and processing in both the

awareness support system and the supported groupware.

- Access to a collaborating person's context in a storyboard-like multimedia representation with temporal dimension, conveying a condensed view of recent activity.

In the remainder of this paper we will first discuss related work. This is followed by an introduction to our new approach to awareness support. Another section describes implementation of a prototype, and provides a scenario-based demonstration of the system. We then discuss insights gained at this early stage of system development, and conclude with an outlook on further work.

RELATED WORK

A variety of methods have been proposed in the CSCW community to support awareness in distributed collaboration. These systems can be classified into three groups, based on the kind of context information they use for promotion of awareness.

- Formal Awareness: Systems monitoring interaction with the computer systems used in the collaborative work environment.

- Awareness of Specific Context: Systems that obtain specific context information of limited scope from specialized sensors such as location sensors.

- Video-mediated Awareness: Systems capturing general context information with cameras to be mediated as black box context.

Formal Awareness

Systems supporting formal awareness monitor interaction of people with their local computer system and distribute this information to the collaborating group. There is a variety of techniques to obtain formal context information. These include tracking of keyboard events, of pointing device position and movement, and of mouse button events [4,6]. The obtained information are white box context, if the promoted events have a meaning within the distributed groupware system, for instance pointer movement in a shared whiteboard. It may also be black box context, for instance if monitored interaction is taken as cue for availability of a colleague.

Awareness of Specific Context

A range of systems have been described that support awareness based on specific context that is captured with specialized sensors, installed in the work environment or worn by collaborating people. In the Active Badge system, the work environment is equipped with sensors that keep track of the electronic badges people wear [15]. This information is available to the entire group and can be used to locate people for collaboration, or to check their availability. Even more specific is the awareness information provided by the Hummingbird, a wearable appliance that hums whenever another Hummingbird is nearby, to help initiate collaboration between their wearers [7].

In the ambientROOM system, the work environment is equipped with further special-purpose sensors to monitor for example whether the telephone is engaged, whether the door is open or shut and other highly specific context information [9]. In the Ambient Telepresence system, everyday objects are equipped with sensors to obtain context information for colleague awareness. For instance, temperature and acceleration sensors are build into the base of coffee cups to track the state of the cups (temperature, location), their manipulation (drinking from cup, playing with cup), and their proximity to other cups [1,3].

The use of sensors to obtain specific context information in the work environment imposes demands on infrastructure but with recent advances in sensor technologies and embedded systems cost is decreasing rapidly. The data collected is highly specific in contrast to information obtained with video but with use of different sensor a rich context information may be acquired. In contrast to video, collected information is usually cheap to process and to interpret due to their specificity. For example, in the Ambient Telepresence system, interpretation of whether a cup is handled for drinking or for playing is done by a small program running on a PIC micro-controller embedded in the base of the cup [3].

Video-mediated Awareness

Formal awareness and awareness based on specific context provide useful context for collaborative work. However they are generally not sufficient to provide distributed collaborating people with a true sense of working together, as they do not mediate cues that people rely on in face-to-face collaboration [4,5]. Such cues are for example obtained from observing facial expression, body language and tone of voice. An approach to make such cues available in distributed collaboration is to use multimedia communication. Media Spaces for instance provide continuous audio and video communication in addition to formal awareness, to mediate rich context information in remote collaboration [2,13]. There are a number of problems associated with the use of videoconferencing as awareness technique:

- Videoconferencing systems can not capture all the cues that may be useful for collaboration, as they provide a static and restricted view of the remote location [4].

- Videoconferencing supports only simultaneous collaborative work, and does not provide context information for participants who work at different times.

- The use of continuous video compromises the privacy of the participants. It is difficult to control the content that is mediated in video streams, and both collaboration-related and non-related information will be shared. Participants may feel that they are watched all the time.

- Moreover, videoconferencing requires high bandwidth, adding to the cost of the collaborative process.

These issues have been addressed in a variety of ways. Instead of sharing the whole video stream, a short period of video [14] or snapshots [10,11] are used to give people a glimpse of the workspace, while reducing required bandwidth. To preserve the participants' privacy, the images in transmitted visual material are masked or blurred through image flittering and other processing techniques [8,16].

A NEW METHOD FOR AWARENESS SUPPORT

Formal awareness, specific context awareness, and video-mediated context are different approaches to collect context information in the workplace and to communicate this information to distributed collaboration partners. While various techniques and systems have been introduced for each of these approaches, we are not aware of work combining different approaches. In particular, we are not aware of work on fusion of video-mediated information with information obtained from specific environment sensors. However many techniques, especially those developed in the field of multi-modal interface and context awareness, show that communication can be improved by fusing information from multiple perception channels [12].

In this paper, we present a new method to support awareness based on fusion of context information from different sources in the work environment. A key concept in this method is the combination of both white box context, interpretable by the system, and black box context, interpretable only by humans. A range of perception methods are integrated in our method to obtain white box context, which in turn is used to process black box context, and to generate condensed awareness information to be promoted to collaboration participants.

The architecture of our method is shown in figure 1. Data collection is based on integration of different sources of context information, in particular integration of audio and video sources with more specific environment sensors and with logical sensors that capture formal context. The collected data is subjected to a processing stage that applies different perception techniques to obtain white box context information from the multimedia data. The white box context information is used for further reasoning about context, and for generation of a storyboard-like representation of context information. The "storyboard" serves as condensed view of recent activity in the workplace, and is accessible for collaboration participants via the network.

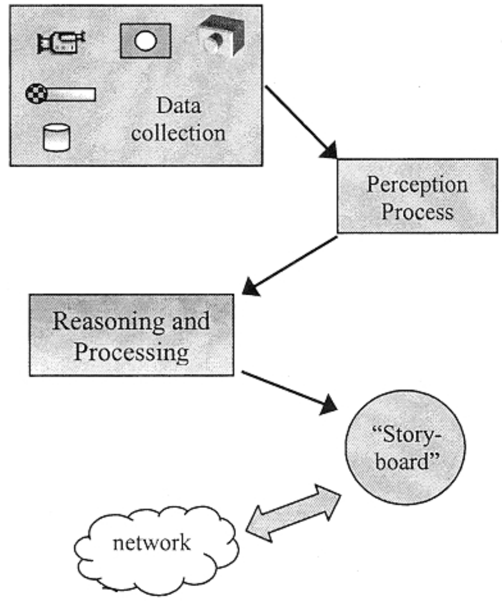


Fig. 1. Architecture of the collaboration log system

Local data collection

Context information in the workplace is collected from different sources. Logical sensors are used to collect infomration on interaction with the computer system, and groupware application. Simple environment sensors are installed in the workplace to sense movement of people, changes in the lighting, status of the telephones and so on. Finally, several cameras and microphones are installed to capture information rich audio and video material from different perspectives in the work environment.

Perception process

Data collection form multiple sources yields a large amount of potentially useful context information. Further processing is required to reduce the amount of data with respect to storage and transmission cost, to filter data that is less meaningful for collaboration, to mask some of the data for protection of privacy, and to abstract higher-level context information from low-level data for further reasoning. In our method, we use a variety of perception methods to filter information, and to obtain information for further reasoning and processing. With these perception methods, meaningful events in the work environment can be recognized and described in a well-defined structure, for instance:

- The name of the contexts: body tracking video
- Start time: 1999/03/21/10:45:31
- Place: conversation place

- End time: 1999/03/21/10:45:34
- Dimension of the description: 3
- Size of each dimension:......
- The data of the description: ......

The collection of meaningful events constitute a record of activity perceived in the workplace. We refer to these events as context. The recognition of the contexts in our system is based on multi-sensor fusion and artificial intelligence techniques.

Motion detection and tracking
We use video analysis in combination with information from passive infrared sensors to detect and track motion of the objects and people in the workspace. Movement in the workspace can be a rich source of context information such as location of the people, whether the door is being opened, and whether a piece of paper is flying to the floor with the wind. Based on the recognition of moving people or objects from video and infrared data, the tracking method can work on the limited objects' movements in static background.

Human face detection and tracking
Human face detection and tracking techniques are applied to locate faces, and to keep track of their location in a scene. This can be used to reducing a video stream to a face stream that contains important cues such as facial expression, and what a person is looking at. To detect and track faces, video information is fused with location information, and image recognition and video analysis are applied on the video material.

Sound analysis
Some kinds of noise or sound have special meaning in the workplace. For example, if the telephone is ringing, it means a phone call is coming in. If there is a loud noise in conjunction with the door being closed, it may mean that was banged shut aggressively. Also sound events such a loud bus driving past the open window constitute useful context, as they may affect the attention of the person in the workspace.

Human voice capturing and background noise analysis
The human voice is obviously a rich source of context information. To capture it for further processing, background noise is analyzed, and filters are build that separate human voice from background noise.

Lighting condition analysis
Analysis of lighting conditions provides immediate cues such as whether artificial light is on, but also provides cues for image analysis and video processing.

Reasoning and Processing

In daily life, human beings use context information to understand situations, and to decide about their own behavior. In the research field of context awareness, many techniques are being developed to provide computer systems with a similar ability of understanding context. However the ability of computers to recognize context in complex information such as speech, gestures, facial expression, and body language is very limited while such context is easily accessible for humans.

To apply context awareness techniques in our system, we introduce the white & black box context mechanism. The core of the white & black box context mechanism consists of a context database, a decision unit using reasoning, and a processing unit as shown in figure 2.

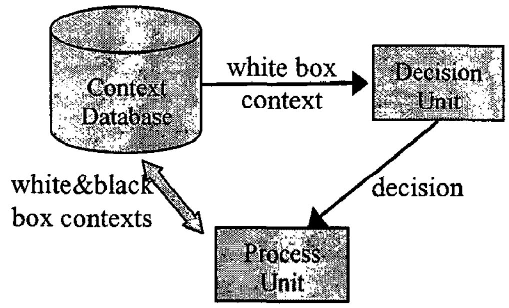


Fig. 2. white & black box context mechanisim

The context database contains the context information obtained in the perception process described above. For each collaborative task, processes and the related reasoning rules are defined on the description values of the possible contexts. For example, we can define: if "the user's face is available", then when "he is asking question to other participants", record his "face expression". All the reasoning rules are stored in the decision unit, which decides depending on recorded context, what kind of processing to apply. These contexts are classified into white box context and black box context according to their description space and processing definitions.

If all the possible states of the description space of a context are well defined with proper processes, it is regarded as a white box context. In the last example, the context "the user's face is available" is a white box context, because its description space is binary, and we can easily define processing for each state. The context "face expression" is regarded as a black box context, because its description space is too complex to define processes for all the possible values. In this case, we assume that the exact meaning of the black box context can only be understood by humans. The white and black box contexts play different roles in the generation of a storyboard for promotion of awareness.

White box context
Only the white box contexts can be used as conditions in the decision unit. They are the material for reasoning. In the processing unit, a white box context can be deleted, or combined with other contexts to produce new contexts. For

example, consider the following white box contexts in the context database:

- "The door is being opened."
- "Motion is detected with Infra sensor near the door."
- "Motion is detected with the camera to the door after filtering out the motion of the door's opening."

In the decision unit, possible decisions for processing of these three white box contexts may for example be:

- Delete all the three contexts.
- Produce the white box context: "person is entering the room."
- Produce the white box context: "the door is opened."

The new contexts are also stored in the context database. We organize the white box contexts within three groups in the context database.

1. Context related to people's location, which includes for instance "people entering the room", "people leaving the room", "there are people in the room", "people in front of the monitor", "people in the conversation place" and so on.

2. Context related to the state of objects, which includes for example contexts such as "the door is open", "the door is shut", "the phone is ringing", "the light condition", and so on.

3. Context related to the people's attention. This group includes for instance "facing to the screen", "telephone conversation", "face to face conversation", "key board typing", "listening to the radio", "opening shared document", "mouse movement", and so on.

Fig. 3. Video/audio streams of the whole scene

Black box context

The black box contexts can not be used in the reasoning based decision unit because of their complexity. However in the processing unit, a black box context can not only be deleted, and combined with other contexts to produce new contexts; it can also be used to extract new contexts. In the example of the last subsection, consider a black box "video stream of the motion in the door view" in the context database. The decisions related to this black box context can for instance be:

- Produce new black box context: "the body video stream in 1.5 seconds".
- Produce new black box context: "the most positive pose snapshot".
- Delete the "video stream of the motion in the door view".

The common attribute of black box contexts is that they are difficult to be understood by computers. A black box context can be "video/audio streams of the whole scene", "user's voice", "body stream", "face stream", "video/audio streams of a conversation", and so on (see figures 3 and 4).

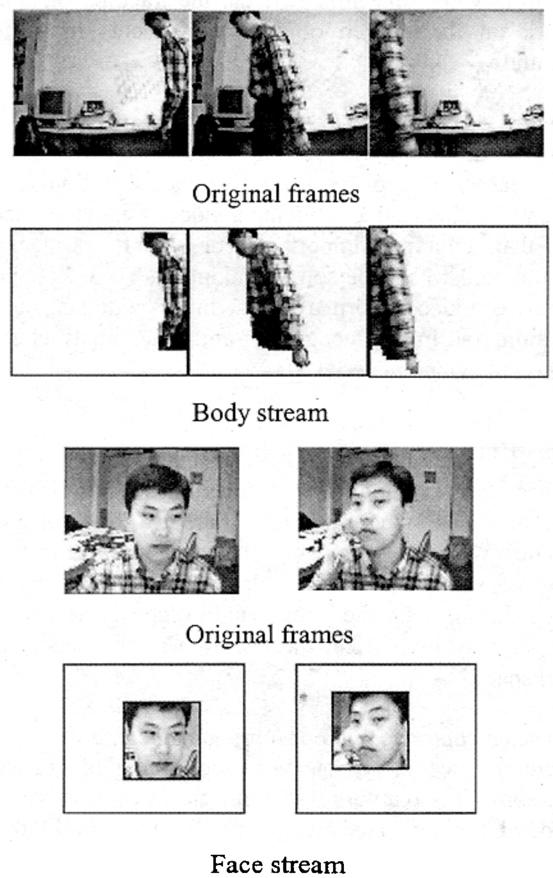


Fig. 4. More complex black box contexts

Because the extraction of new contexts from black box context is mostly based on the white box context and video/audio analysis, the attributes of the same black box context may be different in different groupware system environments. For example, if system A has the ability to detect and tracking the human face, but system B has not. For the same context "video stream of motion in front of the monitor", system A can extract the context "face expression video stream". It is only a part description of the context "video stream of motion in front of the monitor". System B can only use the whole context "video stream of motion in front of the monitor" as building block for awareness information.

The black box contexts can be regarded as the increments of the white box contexts. Because computers lack the perceptual intelligence of humans, the white box contexts can not cover all the useful collaboration cues of the participant's work. The uncovered part of cues is kept as black box contexts, which will be shared to other participants when it is necessary. From another point of view, the white box contexts constrain the content of black box contexts within certain arrangements. This will not only reduce the data volume of the black box contexts, but also enable the system to control privacy of the participant.

Storyboard

Instead of glimpses or glances, which are used as metaphors in real time video-based awareness, we propose to promote awareness in a form similar to storyboards or comic strips. The idea is to provide a condensed view or log of recent activity, rather than a real time snapshot. Continuous exposure to cues via video is replaced by access to a view generated from recorded cues. This approach compromises peripheral awareness but has the advantage of selective access to context, of bandwidth savings, and of applicability in non-simultaneous collaboration. Beyond conventional storyboards, the ones generated by our system are multimedia, content based, and dynamic.

First, the storyboard is multimedia, combining images, text and audio as presentation media.

Second, the storyboard is not fully structured but organized through content-based methods. The pictures and the audio streams are stored in the form of black box contexts. The black box contexts only appear in the storyboard with additional textual explanation. The text explanations are stored originally in the form of white box contexts. These white box contexts can be changed into the text explanations with a well-structured language. We limited the text explanation as a kind of well-structured language, because this property allows other participants to access the log with the index of white box contexts.

Third, the storyboard is produced dynamically, according to by whom it is accessed. Usually, different tasks require for sharing of different contexts among the participants. To make the log for a certain collaborative task, there should be a sharing control to each participant for the contents in the log. For example, for participant A, certain frames in the "face stream" and the audio of a telephone conversation may be put onto the storyboard, while for participant B only a text explanation appears: "Taking a phone call!".

IMPLEMENTATION & DEMONSTRATION

We have implemented an experimental system for which we equipped an office room with three cameras, microphones, Infrared sensors, and other environment sensors. All these devices are used to collect data from the user and his work environment. The context awareness and white & black box context mechanism recorded the events in this work environment under the assumption that the total number of the users and visitors are no more than three. The storyboards are written into HTML documents with standard image and audio format. The participants can access it easily through common web browsers.

To demonstrate the operation of the system, a typical working morning for a working group is presented. In this group, participant A (PA) and participant C (PC) share the same office. The project manager (PB) and participant D (PD) are only accessible through telecommunication. According to the profiles for each other participant, a series of storyboards of PA are produced for the other participants.

Available for: PB, PC, PD

14.01.1999, 8:00
No person in the office

Available for: PB, PC, PD

The room is empty

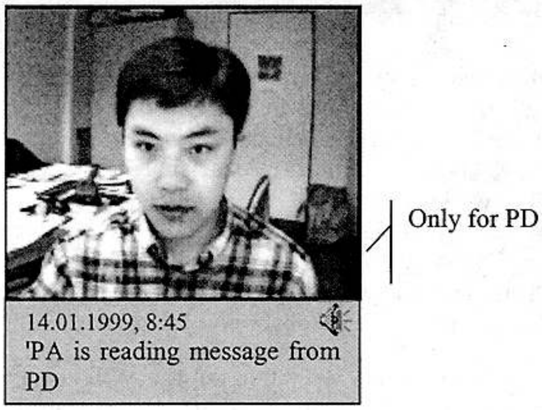
14.01.1999, 8:30
Person enters the office
Login as PA

Available for: PB, PC, PD

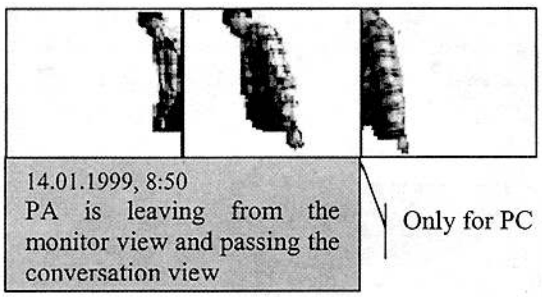
PA comes into the room

There are three messages for PA. Some of them connect to the storyboards of other participants.

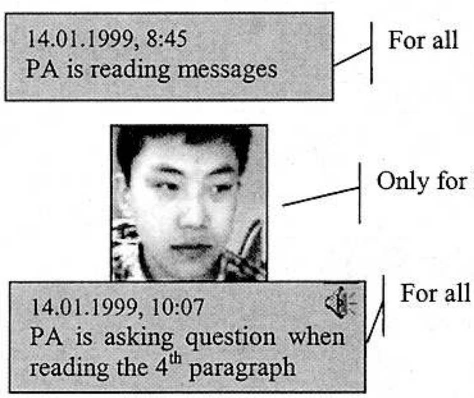
- PC worked deep into the night yesterday and is not available. He had a video message for you.
- PD has a message.
- There is a message from PA's wife.



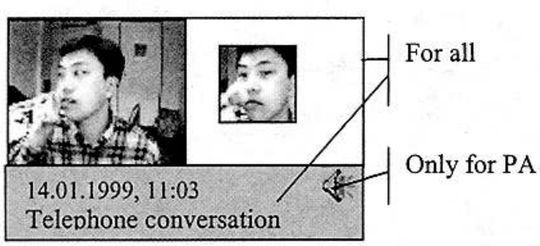
PA begins with the PD's message, and then, his wife's message……



Then, PC's video message is displayed on the screen. It tells PA that he left a device in a box in the corner of the room. PA fetches the device from that box, which is recorded in the storyboard for PC.



After reading his messages, PA has to review instructions for a new system maintained by another participant of this project. After reading four paragraphs, PA finds a new term in the next sentence. He selects the "question" button, and



the result shows that the only available participant is PB, because others have not yet read these instructions. While taking a phone call, PA finds a message displayed on the screen. It says that PB will answer his question, and how soon he can finish his conversation.

Figures 5 and 6 show the different storyboards generated for participants B and C.

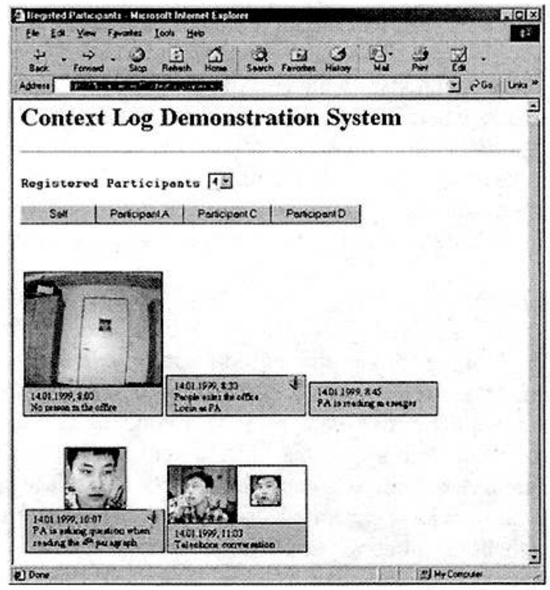


Figure 5. The storyboard of PA generated for PB

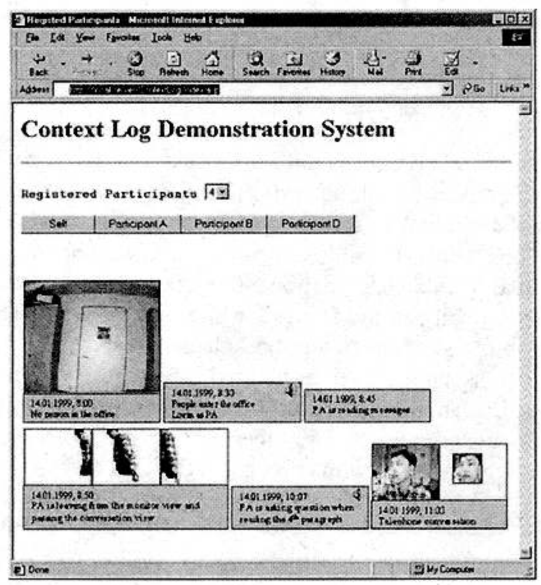


Figure 6. the storyboard of PA for PC access

DISCUSSION OF STORYBOARD-BASED AWARENESS

The storyboards as shown in figure 5 and 6 provides the collaboration partners of particpant A with a condensed view of recent activity in his work environment. It contains useful cues for collaboration at a glance. Participant B for example can see at a glance that his help is required but that PA is currently on the phone. The storyboard for PC contains different cues based on the different collaborative task PA and PC are engaged in. The described storyboards also illustrate how privacy concerns can be addressed, for example the context "reading his wife's message" does not appear in the storyboards, and when A walks passed the conversation view, only his movement is shown but not the background.

The storyboard-like representation provides collaboration participants with access to a history of events and cues. For initiation of collaboration it will often be useful to have access to past cues and not just to current context. Further, the recording of cues in a storyboard enables awareness support for people working non-simultaneously.

Fundamental for storyboard-based awareness support is the combination of white box and black box context. White box context is the backbone of the storyboards, supporting structured representation, content-based access, and reasoning. Black box context is additional information, added to convey cues that the system itself can not recognize. Reasoning based on white box context is used for protection of privacy, and for dynamic selection of cues relevant for a given collaborative situation.

The material for the storyboards is produced fully automatically. Perception methods give the system the ability to extract context information from audio and video streams, and to record context in a well-structured way for further processing and reasoning.

However, most context awareness techniques used in our system are based on perception methods and semantic analysis approaches. The performance of these approaches limit the efficiency of the awareness support system. Applicability of semantic analysis approaches depend on the complexities of the scene. In our system, the anaylized work environment is limited to a single room and one or two people. The interactions of the people and the events in the environment are also limited as reported in the previous section.

CONCLUSION

The storyboard-based system proposed in this paper contributes new concepts for awareness support. First, it promotes fusion of information from different sources in the work environment, in particular fusion of specific context obtained from sensor infrastructures with general context as captured in audio and video material. Secondly, it introduces the white & black box mechanism, to combine machine-interpretable structured context information with multimedia information that carries detail and subtle cues for human interpretation. Thirdly, to obtain structured context information a variety of perception are integrated in the system. And finally, a storyboard-like collection of cues is produced to promote awareness of recent activity in a workplace. The storyboard is multimedia, content-based and dynamic.

An experimental system was implemented to demonstrate these ideas and concepts. It was restricted to a single office room but demonstrated automatic production of storyboard content from a number of video streams and environment sensors. It also served for investigation of scenarios, considering the use of reasoning for compilation of storyboards adapted to different collaborative tasks and privacy requirements. However, the experimental system also showed limitations imposed by perception techniques.

The early work presented was primarily concerned with overall architecture of the approach, and with methods for automatic production of storyboard content. Future work will be more focussed on the use of reasoning. For example, while it is clear that reasoning can be used to protect privacy, it is not clear how transparent this should be for the user. This is but one usability issue that we need to investigate with an extended system supporting studies of how the system affects the users and collaborative task performance.

REFERENCES

1. Beigl, M. and Gellersen,H.W. Ambient Telepresence: Colleague Awareness in a Smart Environment. In Managing Interactions in Smart Environments, Proceedings of MANSE'99, Dublin, Ireland, 13-14 December, LNCS, Springer-Verlag Heidelberg, 1999.

2. Bly, S., Harrison, S. and Irvin, S., "Media spaces: Bringing people together in a video, audio, and computing environment", Communications of the ACM 36(1), p: 28-46, 1993.

3. Gellersen, H.W., Beigl, M., and Krull, H. The MediaCup: Awareness Technology Embedded in an Everyday Object. In Gellersen, H.W. (Ed.) Handheld and Ubiquitous Computing, Proceedings of HUC'99, Karlsruhe, Germany, 27-29 September, LNCS No. 1707, Springer-Verlag Heidelberg 1999, p. 308-310.

4. Greenberg, S., "Real Time Distributed Collaboration", http://www.cpsc.ucalgary.ca/grouplab/papers/

5. Gutwin, C., Roseman, M. and Greenberg, S., "A Usability Study of Awareness Widgets in a Shared Workspace Groupware System", in Proceedings of ACM Conference on Supported Cooperative Work, p: 16-20, 1996.

6. Gutwin, C. and Greenberg, S., "Workspace Awareness", Position paper for the ACM CHI'97

Workshop on Awareness in Collaborative Systems, p: 22-27, 1997.

7. Holmquist, L.E., Falk, J. and Wigström, J. Supporting Group Collaboration with Interpersonal Awareness Devices. In Personal Technologies, Vol. 3, No. 1&2, 1999, p.13-21.

8. Hudson, S. & Smith, I., "Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems", in Proceedings of the ACM conference on Computer Supported Cooperative Work, p: 248-257, 1996.

9. Ishii, H., Wisneski, C., Brave, S., Dahley, A., Grbet, M., Ullmer, B. and :Yarin, P., "ambientROOM: Integrating ambient media with architectural space", in *CHI'98 Video Program*, 1998.

10. Johnson, B. and Greenberg, S., "Judging People's Availability for Interaction from Video Snapshots", in Proceedings of the Hawaii International Conference on System Sciences, 1999.

11. Lee, A., Schlueter, K. and Girgensohn, A., "NYNEX Portholes: Initial User Reactions and Redesign Implications", in proceedings of ACM SIGGROUP conference on Supporting Group Work, p: 16-19, 1997.

12. MacLeod, B. E. F. & Summerfield, A. Q., "Quantifying the contribution of vision to speech perception in noise", British J. of Audiology, Vol. 21, p: 131-141, 1987.

13. Mantei, M., Baecker, R., Sellen, A., Buxton, W, Milligan, T. and Wellman, B., "Experiences in the Use of a Media Space Remote Synchronous Collaboration", in proceedings of ACM conference on Human Factors Computing Systems, p: 203-208; 1991.

14. Tang, J.C., Isaacs, E., and Rua, M., "Supporting Distributed Groups with a Montage of Lightweight Interactions", in Proceedings of the ACM conference on Computer-Supported Cooperative Work, p: 23-34, 1994.

15. Want R., Hopper A., Falcao V., Gibbons J., "The Active Badge Location System", ACM Transactions on Information Systems, Vol 10, No 1, 1992.

16. Zhao, Q. A. & John T. S., "Evaluating Image Filtering Based Techniques in Media Space Applications", in Proceedings of the ACM conference on Computer-Supported Cooperative Work, p: 11-18, 1998.