

Cognition Based Collaboration Log for CSCW

Datong Chen & Hans-Werner Gellersen

TecO (Telecooperation Office), University of Karlsruhe

{hwg}@teco.uni-karlsruhe.de

Abstract:

This paper presents a method to promote awareness support system in both real time and non-real time computer supported collaborative work. Analysis and recognition techniques are applied to the collected information and to obtain contexts from formal computer input data, simple monitoring data, and audio/video-based data from the local workspace. The contexts information are interpreted into white box contexts and black box contexts and stored in a special designed content-based multimedia database called collaboration log. The collaboration log records a long period of working in the local workspace, which can be further composed and shared to other participants by the awareness support system. By using this cognition based collaboration log, the system can support the awareness of the workspace through both current and history information, while preserve the private information automatically and efficiently.

Keywords: collaboration awareness, cognition, groupware, content-based multimedia

Introduction

In face-to-face collaboration people may use individual cues to stay aware of other people's availability for interaction, and to assess whether it is appropriate to initiate a specific collaborative activity. Participants can use the 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 with a collaborative task. Comparing with the traditional face-to-face collaboration, in distributed collaboration, people have only limited information on collaboration context available to take decisions that determine the collaborative workflow. This problem is addressed by awareness support systems, designed to provide collaborating people with additional informal context information, especially the audio/video context information.

A large variety of awareness support systems have been discussed. In addition to formal awareness, these systems promote awareness of informal context. The

general approach is to continuously or periodically capture and transmit context information in real time, which can support the workspace awareness in real time collaborative work. Most systems capture and transmit context information that is not interpreted by the system. For such context information we use the term *black box context*. For example, media spaces capture and transmit audio/video as black box context from one location to another. In contrast, other awareness systems interpret the collected context information to obtain and transmit, which we call *white box context*. For example, active badge systems transmit locational information as white box context, suited for further processing in a groupware system.

In this paper, we propose a method to apply analysis and recognition techniques to improve the awareness support system. The approach is to interpret the formal, sensor-based and audio/video data in a local workspace into both black box and white box context and keep a history of both of these formal and informal collaborative context information to support further awareness. A host of recognition techniques are used to derive white box context from collected data obtained from both simple monitoring techniques and audio/video capture. According to the derived white box contexts, the collected data is organized into a special designed content-based multimedia database [15] as contexts, which is called collaboration log. In contrast to the common approach, context information is not transmitted continuously but accessible in certain forms and allows people look into the collaboration history at any time.

Related Work

A variety of methods have been developed in the CSCW community to support awareness in distributed collaboration. Formal context monitoring groupware systems share formal awareness information, which represent the people's interaction on shared groupware process, simply through tracking the process launching, keyboard events, mouse movement, and button hits [1, 2]. Some groupware systems employ simple sensors, which are fixed into the workspace or mobile with the people, to obtain context information. For example, in the *ambientROOM* information about the motion and

the state of telephone, door, and other equipment is collected with simple sensors [3]. In the Active Badge system, collaborating people wear small devices with wireless communication ability to obtain and transmit information for location awareness [4]. In the smart badge this concept is extended to obtain further context from people wearing the badges: “it knows when you are hot and sweaty” [5]. Both the formal and sensor-based approach to awareness imposes demands on infrastructure but the data collected can often be interpreted with little processing, as signals are stable and easy to structure.

However, the formal and sensor-based awareness can offer important context for collaborative work, but they are not satisfactory for providing participants with a feeling of working together. Greenberg [1, 11] has discussed the reason that people’s activities rely on many subtle cues from voice, body language, and so on, which are available in face-to-face collaboration but which lack in distributed collaboration. Based on multimedia communication technologies, the cooperative cues from the participants’ behavior and the working environment can be partly offered through video and audio streams, for example, the videoconference technique [6, 14]. Unfortunately, there are many disadvantages when videoconferencing is used as awareness technique:

- First, some researchers find that videoconference can not capture all the useful information [1]. Many information carried by events in the working environment are omitted.
- Second, the videoconference supports only the simultaneous cooperative process. When participants work at different time, the informal collaboration information will be not available.
- Third, it threatens the privacy of the participants. The contents of the videoconference are difficult to be controlled. Both cooperation related and non-related information will be shared through the videoconference. It makes the participants feel that they are monitored at any time.
- Moreover, videoconference asks for a very high transmitting bandwidth connection between participants that limits the audio/video information sharing through low bandwidth Internet connection and increases the cost of the cooperation process.

Many approaches have been discussed to address these issues. Instead of sharing the whole video stream, short periods of video [7] or snapshots [12, 13] are used to give people a glimpse of the workspace. In this way, the bandwidth cost is reduced sharply. A lot of methods were also presented to preserve the participants’ privacy. In these systems, parts of the visual material in the sharing images are masked or blurred by using image flittering, image differencing and other processing techniques [8, 9].

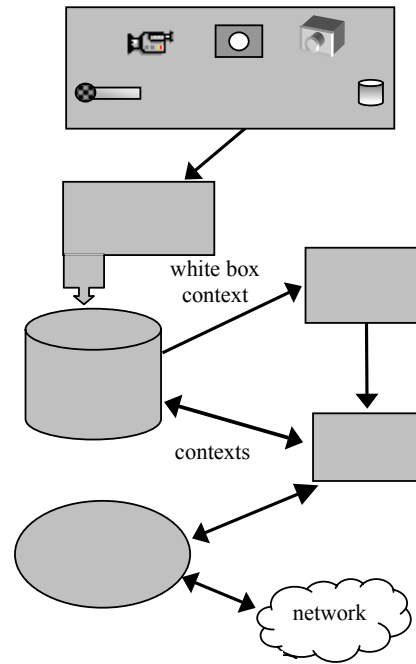


Figure 1. Architecture of the collaboration log system

Cognition based workspace awareness

The idea of our approach is to use analysis and recognition techniques to interpret the data collected from the individual workspace, select and store the useful materials from the data to support awareness of this workspace. The overall architecture of this awareness support system is shown in figure 1. The system collects the data from the computer, the simple sensors, video cameras and microphones distributed in the workspace. The raw data are regarded as black box contexts and are used to derive the white box context in the perception process. The recognition algorithms used in the perception process include the translation of the formal contexts, simple processes for the sensors, audio/video analysis and fusion methods based on multiple percept channels. Both the white box and black box contexts are stored into the collaboration log at first, and then refined by the predefined rules in decision unit and process unit. The access of the collaboration log is controlled by a communication agent, which mediates the communications between participants according to the context information. Cognition techniques play an important roll in running the system automatically, and will be discussed as an emphasis in this paper.

Black box context

If not all the attributes of a context can be interpreted by the system, this context will be regarded as a black box context. The raw data, for example, “the video stream of the whole view of the office”, “the audio data from one

of the microphones”, are black box contexts obviously. There are also some black box contexts in the results of the recognition process, because it is difficult to build the system with the same intelligence as the human being’s. The context “body video stream when entering the room” is derived from the video stream obtained from a camera faces to the door of the office room. In our system, there are no algorithms to recognize the body gesture of the human being, so that this context will be treated a black box context and shared to other participants with the whole video stream when it is necessary. From other examples of this kind of black box contexts, such as “face expression when reading a certain message”, we can see the importance of this kind of context in supporting awareness.

White box context

If all the attributes of a context can be interpreted by the system, this context will be used as a white box context, for instance, the context “motion of the door is detected” is a white box context with binary value. The white box contexts can be derived from the black box context or deduced from other white box contexts. The awareness support system maps every white box context to known states, and uses them to control the production of other new white and black box contexts.

Cognition based context derivation

The information of the workspace can be kept by recording all the scenes and events happened in this workspace. However, in this way, the record data grow very fast and are difficult to be stored and transmitted. Moreover, if these raw data are shared directly, it costs too much time for the other participants to browser them. In the process of real collaborative working, the remote participants are interested in only certain kind of information that is carried by only a part of events in the workspace. For example, they may hope to see the video stream of someone when he is asking a question, but not the video of an empty room. On the other hand, the local participants prefer to only sharing necessary part of information to others, but preserving some private data from them. In order to efficiently organize and share the local data to support the awareness of the individual workspace, plenty of cognition techniques are employed to filter out the less meaningful part of the data and to recognize the private sensitive part of the data. After interpreted by the perception processes, the data are represented with well-defined structures, which are called contexts, 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 algorithms of the cognition are selected depend on the properties of the local workspace and the

collaborative task. They can be statistic methods, pattern recognition, audio/video processing, and multi-sensor fusion approaches.

Motion detection and tracking

The motion of the objects in the workspace is an essential event to show that the environment is changing. It can be detected and tracking through sensor-based recognition and video analysis.

The acceleration sensors fixed on a rigid object properly measure the motions of this object precisely in specified directions. The motions of the rigid objects, such as the motions of door and the windows, are easy to be obtained with the acceleration sensors.

The motion of the human being can be detected and roughly tracking by using the passive infrared sensors. Figure 2 shows the output of one passive infrared sensor when a person is closing and leaving it.

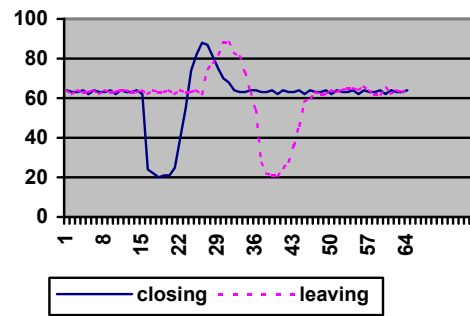


Figure 2. Motion tracking with a passive infrared sensor

The distance between moving person and the passive infrared sensor can be measured roughly with the following equation.

$$DIS(t) = \left| \Omega - C_1 \sum_{i=0}^t (x_i - \theta) - C_2 (t - t_0 + 1)(t + t_0) \sum_{j=0}^t (x_j - \theta) \right|$$

Here, C_1 , C_2 , θ are system calibration constants, Ω is the maxim detect distance of the sensor.

Region-based video analysis is used to extract the motion information from the visual data. The algorithm assumes that the camera is fixed and the background in the scene is static. At first, the algorithm computes the difference image from the current image and a stored background image. And then, the difference image is filtered with a region-based image filter to obtain the valid regions. If the valid region occupies most part of the image, the algorithm will update the background image with the current image. Otherwise, the algorithm will output the mask of the moving region.

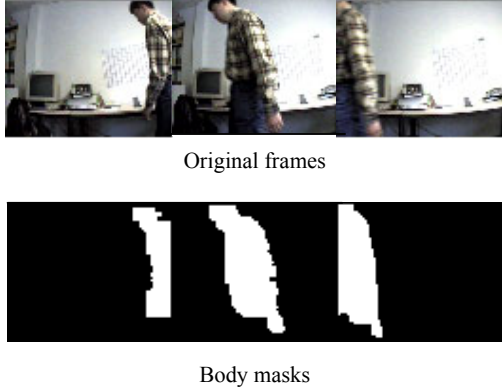


Figure 4. Video-based motion detection and tracking

In certain cases, the objects in the background are not always static, for instance there is a door in the view, the background image still can be estimated by using the motion information from other sensors. The positions of the door can be located from the data of acceleration sensor, and then the region covered by the door can be filter out. Combining with other sensors, the video-based algorithm works stable when there are no more than two persons in the workspace.

Human face detection and tracking

After knowing where the person is, the system tries to locate the human face and to keep track of his location in a scene. The human face expressions bring the rich information to the other participants and the location of the face may show where the person is looking at roughly. The human face detection and tracking algorithm bases on the template-matching algorithm. The template consists of 9 regions, which is showed in figure 5.



Figure 5. Face template

The nine regions $\tau_0, \tau_1, \dots, \tau_8$ represent eyebrows eyes mouth, cheek and hair. We use the average grey value of the pixels in these regions as the feature vector for the human face.

$$a_i = \frac{1}{\text{sizeof}(\tau_i)} \cdot \sum_{(j,k) \in \tau_i} f(j,k)$$

$$F = (a_0, a_1, \dots, a_8)$$

If the distance of the current feature vector F with one of the standard face feature vector is smaller than certain threshold, the algorithm will mark this area as a human face.



Figure 6. Human face detection and tracking

Figure 6 shows the results of the human face detection and tracking. The algorithm can detect the human face in dynamic background depending on the people location tracking results, but limits the orientation of the human face in $\pm 45^\circ$. In the first and the second images, the human faces are detected. In the third image, our algorithm fails to find a human face because the orientation of the face turns out of the range of 45° . This information is helpful for the system to decide that it is not a good time to display something to this user.

Audio analysis

The audio analysis focuses on recognizing the voice of human being, the background noise and some kinds of sound have special meaning in the workplace. The human voice is an important way for human beings to communicate with others. The investigation in [16] indicated the fact that the human voice is even more important than the video information in the distributed collaborative work. The background noise in certain office room takes on some certain patterns. When there is no artificial audio device opening in the room, the background noise patterns are rather stable. The recognition result of the background noise is not only helpful to detect human voice and fade the noise when record the human voice, but also is valuable to detect other sound related events. For example, if the telephone is ringing, it means a phone call is coming in; if there a strange loud sound, it may mean a cup fell down to the floor. These special sound events affect the attention of the people in the workspace, and mediate their interactions with the system in some cases. The audio analysis algorithm extracts the features from both the time and frequency domain.

Lighting condition analysis

Several simple sensors are distributed and fixed in certain location or on important objects' surface in the local workspace to measure the light conditions of the specified area. Through analyzing of these light sensors, we can obtain the useful contexts about the information of the flashing of the artificial lights and shading caused by the motions of clouds or human being, and so on. These contexts are used to activate the video analysis algorithm, filter the invalid moving detected regions, improve the quality of the video stream and adjust the parameters used in video analysis algorithm.

Reasoning, processing and communication

The recognized contexts are the common descriptions to the local workspace. According to the demands of different tasks, these contexts need to be refined in the collaboration log with context awareness. Context awareness is a kind of technique that people use the computer to simulate part of people's context processing ability to build intelligent decision making system. In our system, the recognized contexts are used to deduce new contexts with rule-based reasoning. For certain task, the rules can be predefined by the user himself.

The refine work of the collaboration log involves mainly three kinds of processes. First, the processes focus on local contexts, include producing new contexts, delete the invalid contexts, etc. Second, the contexts obtained from remote collaborative participant are processed with the local contexts together. Some new contexts may be produced to respond certain remote contexts, and others may be used to carry the feedback from the remote participants. The last kind of the process is to build the dependencies between the different contexts and also between the contexts and certain processes. For example, the context "ask a question" should depend on the context "first available participant", so that the system can share the former context to the exact people when the later context appears.

The communications between distributed participants are controlled by a specially designed agent. The agent decides the time, address, formations and the contents of each communication according to the both local and remote contexts.

Experiment and results

The participant's workplace in the experiments is limited in a single office room, which equipped with cameras, microphones, passive infra sensors, acceleration sensors, monitor, speaker, light sensors and the wireless communication system. All these devices are used to collect the data from both the people and the environment, then send the data as digital form to a PC through cables or wireless connections.

One of the scenarios is that, the room was empty until the participant A comes into the room. After he sits in front of the monitor and logs into the system, there displays the message list:

- Participant B came at 8:00 o'clock, and are still working now.
- Participant C worked deep into the night yesterday and is not available. He had a video message for you.
- Participant D has a message.
- His wife has a message.

He begins with the Participant D's message, and then, his wife's message. The participant C's video message is displayed on the screen. It tells participant A that he had put a device into the box at the Corner of the room. He fetches the device from that box. "A" fetches the device from the corner. This morning, participant A should review instructions about a new system maintained by another participant of this project. After reading four paragraphs, Participant A finds a new term in the next sentence. He points the "question" button, and the result shows that the only available participant is participant B, because others have not read this instructions. When taking a phone call, Participant A finds a message display on the screen. It says that participant B will answer his question, and how soon he can finish his conversation. "1 minutes" Participant A replies.....

The collaboration log of participant A is produced as the following description. Here we use the only the name of to represent the whole description of the contexts.

- 14.01.1999 AM 8:30: "the first people entering the room" {"body video period" "audio period"}+other attributes;
- 14.01.1999 AM 8:40: "people logon as the user A";
- 14.01.1999 AM 8:45: "reading message No. ##### from participant D" {"body video period" "audio period"}<- "depend on: message No. ##### checking";
- 14.01.1999 AM 8:47: "watching participant C's message" {"face expression period"}=>"body tracking video stream"+other attributes;
- 14.01.1999 AM 10:32: "ask question when reading the 4th paragraph" {"face expression period" "voice"}<- "hanging for participant B's context";
- 14.01.1999 AM 11:03: "telephone conversation" {"face expression period" "body video period"}->"involve the context 14.01.1999 AM 11:05"+other attributes;
- 14.01.1999 AM 11:05: "reply participant B's message" {"face expression period" "body video period" "voice"}+other attributes;

This prototype is implemented within a single office room, which is assumed that the total number of the users and visitors is no more than two. Every context can use only the name and the correspond structure that have already registered in the system. The collaboration log can be accessed as storyboards, which consists of pictures, audio and text explanations and are written into the HTML format documents with standard image and audio format. The participants can access it easily through the common web browsers.

Discussions and conclusions

By accessing this collaboration log, the distributed participants can get a glimpse into the recent history of others. The method discussed in this paper has four merits in building awareness support groupware system. First, with cognition techniques, the groupware system acquires and organizes the awareness support information into a collaboration log automatically. Second, it offers the participants to access the history of others' workspaces that support the awareness of both real time and non-real time computer supported collaborative work. Third, the cognition techniques make it is possible to build log as content-based structure, which not only reduce the bandwidth demand of the system, but also can preserve the privacy freely in the communication. Content-based structure allows the system to produce different contents according to the accesses of different participants. The privacy is protected not through blurring the image and audio, but through only sharing the necessary parts to proper accesses while keeping all the subtle cues of the sharing parts. Finally, the white & black box context mechanism allows the systems with different level of "intelligence" to work together. Some white box contexts in one system can be processed as black box contexts in low-capability systems. The new algorithms can be added in one system to make it smarter without affect others.

However, as we described above, the system can only be used in limited situation, such as single room, no more than two people, and so on. In more complex cases, the system will produce too much black box contexts, which carry plenty of data that can not be stored and shared efficiently. To further improve the performance of the awareness of the workspace, more intelligent analysis and recognition algorithms, which are robust in more complex situations, should be investigated in the future.

References

1. Greenberg, S., "Real Time Distributed Collaboration", <http://www.cpsc.ucalgary.ca/group/lab/papers/>
2. Gutwin, C. and Greenberg, S., "Workspace Awareness", Position paper for the ACM CHI'97 Workshop on Awareness in Collaborative Systems, p: 22-27, 1997.
3. 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.
4. Want R., Hopper A., Falcao V., Gibbons J., "The Active Badge Location System", *ACM Transactions on Information Systems*, Vol 10, No 1, 1992.
5. Beadle, H. W. P., Harper, B., Maguire Jr., G. Q., and Judge, J., "Location Aware Mobile Computing", in

Proceedings of the IEEE/IEE International Conference on Telecommunications, 1997

6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. Koenen, R., Pereira, F., and Chiariglione, L., "MPEG-4: Context and Objectives," *Signal Processing: Image Communication*, Special Issue on MPEG-4, Vol. 9, Nr. 4, May 1997.
16. Watts, L., Monk, A., Daly-Jones, O., "Inter-Personal Awareness and Synchronization: Assessing the Value of Communication Technologies *The Role of Cognitive Science in Human-Computer Interaction*", *international Journal of human-computer studies*, v. 44, p: 849-873, 1996.