Context-Aware Computing for Assistive Meeting System

Peng Dai
Tsinghua National Lab for Information Science & Technology
Tsinghua University, Beijing, China
+86-10-6278-2406
daip02@mails.tsinghua.edu.cn

Guangyou Xu
Tsinghua National Lab for Information Science & Technology
Tsinghua University, Beijing, China
+86-10-6278-2406
xgy-dcs@tsinghua.edu.cn

ABSTRACT

Human-centered computing and implicit human computer interaction will be the future computing models, which can be realized in various pervasive computing environments. Thus computer understanding of human actions and intentions becomes the key. Context plays a significant role in the understanding of multi-party human interactions such as meetings. Dynamic context in this paper cannot be sensed through traditional context-aware approaches; instead a probabilistic framework is required to perform online analysis of multi-level context. Therefore this paper presents a Dynamic Context Model to solve the problem of context awareness toward group meeting analysis and services, which includes multimodal analysis of group interaction scenarios and provision of attentive services to the users. According to our concepts, a distributed multimedia processing system has been implemented in the smart meeting room and preliminary experimental results demonstrate the effectiveness of our approach.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding – *architecture and control structures, video analysis.* I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – *sensor fusion.*

Keywords

Assistive meeting system, context-aware computing, Dynamic Context Model, group meeting analysis.

1. INTRODUCTION

The next generation computing will be about anticipatory user interfaces based on multiple intelligent sensors distributed in the environment, which should be human-centered and operate in the background [1]. The key research issue of human-centered computing is to integrate implicit human computer interaction into our daily lives, i.e. computer system should analyze users' actions and intentions based on multimodal sensor data, and further provide attentive services to users. The computing devices need to be transparent to the users and the services should be non-intrusive. Context awareness plays a significant role in the domain of implicit human computer interaction, since context is tightly correlated to the analysis of human actions and intentions in two aspects. Firstly, appropriate understanding of human behavioral and social signals highly depends on the context, e.g. the same action may convey distinct meanings in different context. Secondly, context has to be considered for picking up the focus of attention in process and fusion of sensory data across multiple modalities [1].

In this paper, the problem of context awareness is defined as online analysis of human actions, intentions and overall situation toward multi-party face-to-face interactions such as group meetings. Dourish and Bellotti [2] has presented early definition of awareness in Computer Supported Cooperative Work (CSCW) systems: "*Awareness is an understanding of the activities of others, which provides a context for your own activity*". In our work, awareness of individual actions and intentions also take the actions of others into account. However in [2] the group users are distributed and interact with each other through manual operations of computers; while in our work users interact with each other face-to-face and no explicit operations of computers are needed in our system. Besides, multimodal information processing and fusion based on the audio and visual sensors adds to the complexity of our work.

Most of the previous works on context-aware computing presented context model to deal with the problems such as context storage, sharing and management in the field of ubiquitous computing [4], which extract contextual information directly from various sensors and are not capable of handling the multi-level context sensing problem we face in this paper. As a recent application instance of ubiquitous computing, Tan et al [5] presented an event-driven context interpretation approach to generate high-level contexts in Semantic Spaces, which mainly dealt with single-user situations and used logic inference for event and context reasoning. However this method cannot solve the multi-party human interaction analysis problems, which involves hierarchical semantic analysis based on multimodal sensor data fusion. In this regard, a probabilistic reasoning model is required for context inference as we did in this paper. Recently some efforts have been made towards the context-aware visual computing in smart environment. Crowley [6] proposed a framework for context-aware observation of human activity. A distributed multimedia system was presented in [7] to capture dynamic context such as changes of users' locations and face orientations in the intelligent environment. However the context models in these literatures did not deal with such composite multi-

PETRA'08, July 15-19, 2008, Athens, Greece.
Copyright 2008 ACM 978-1-60558-067-8... $5.00

party interaction problems which consist of multiple abstract semantic levels as in this paper.

In this paper we aim to solve the context awareness problem so as to achieve online analysis and services during group meetings. There have been some related works on automatic recognition of group events concerning meeting scenarios. The EU research project M4 and its follow-up project AMI mainly dealt with group event analysis based on multimodal meeting corpus [8], [9], [10]. McCowan et al. [8] applied Hidden Markov Models (HMM) for the recognition of group actions in meeting scenarios based on audio-visual information. Zhang et al. [9] extended the work with a two-level HMM framework to model individual and group actions simultaneously. More recently new types of multimodal features such as prosody, speaker turns etc. were used to classify meeting scenarios, and Multistream Dynamic Bayesian Model was adopted in [10]. However, most of the related works in the field of group meeting analysis were constrained to the offline mode and no attentive services were provided to the meeting participants based on the analysis results.

In this paper, implicit human computer interaction is introduced as the essence of human-centered computing. Group meeting as a typical multi-party human interaction scenario is adopted for our research work. Our objective is to make compute system understand current context of group meetings based on the audio-visual information at each time step, and provides intelligent services to the participants instantly. Dynamic Context Model is proposed to tackle context aware problems for group meeting analysis and services, which makes our work have the advantage over previous approaches in that it analyzes multi-level context online. Based on this idea, a distributed multimedia information processing framework is implemented to support context-aware computing in fulfilling meeting analysis and services.

The rest of the paper is organized as follows. Context and context awareness in multi-party human interactions are introduced in Section 2. Dynamic Context Model as the solution for context-aware computing is described in Section 3. Section 4 introduces the implementation framework of our system. Experimental results are presented in Section 5 and conclusions are drawn in Section 6.

2. CONTEXT IN ASSISTIVE MEETING ENVIRONMENT

2.1 Implicit Human Computer Interaction

To make computing devices invisible and serve users better without distracting them from their work or leisure activities, human computer interaction (HCI) needs to evolve from computer-centered designs to human-centered designs [1]. The essence is to introduce implicit human computer interaction into our daily lives. Implicit human computer interaction is a typical HCI model for pervasive computing.

In the concept of implicit HCI, computer systems analyze users' actions and intentions based on multimodal sensor data, and further provide attentive services to the users without drawing their attention away from their current tasks. For instance, in multi-party human interaction scenarios such as group meetings shown in Fig. 1, a lecturer marks on the touch-sensitive screen during presentation can be taken as explicit input for the computer system; however such actions as audience raising hand and asking questions can be regarded as implicit input. Such human actions cannot be categorized as direct commands toward the computer system, but are detected, recognized and explained as implicit input for the system, based on which assistive services are provided to the users in the physical space as implicit output of the computer system. Implicit HCI makes computing devices transparent to users and allow them to focus on their own tasks.

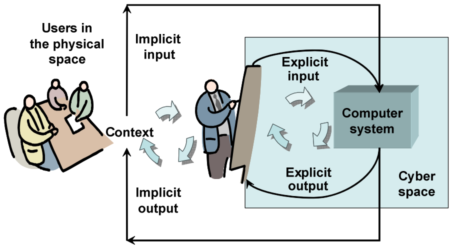


Figure 1. Implicit human computer interaction

The prerequisite of implicit HCI is that computer system can understand users' intentions according to their actions in the physical space so as to generate implicit input. Accurate understanding of human actions and intentions requires the guidance of contextual information, since the same actions might communicate various semantic meanings in different context. Therefore context awareness is the key to implicit HCI.

2.2 Dynamic Context in Meetings

In this paper, we focus on the research of context-aware computing in group meetings, which are typical multi-party human interaction scenarios. Our objective is to set up a smart meeting room and provide intelligent services to the users in an online manner during meetings. The setting of our smart meeting room is illustrated in Fig. 2. Inside the meeting room, three fixed cameras are deployed to monitor the space from different perspectives. Three linear microphone arrays are placed on the meeting table to collect speech information during meetings. A Pan-Tilt-Zoom camera is installed on the table to focus on significant objects.

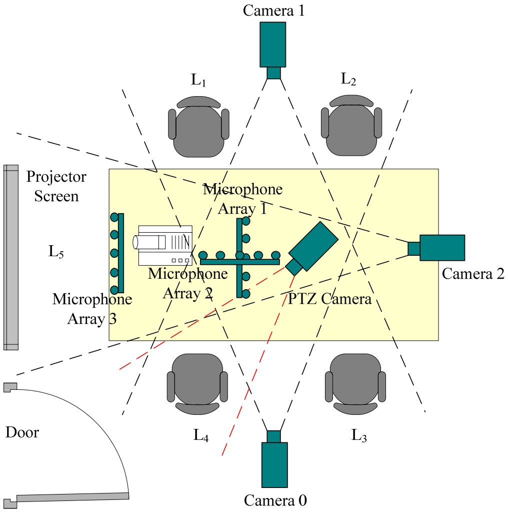


Figure 2. Smart meeting room setting

In traditional desktop computing modes, context of computer system is relatively fixed, i.e. it can be set by users in advance and rarely changes during human computer interactions. However in human-centered computing modes, human computer interaction context is dynamic. For instance, in multi-party human interactions, human states, actions and expressions may change dynamically, and their interaction states will also be varied along with time. Therefore computer system has to sense the dynamic context at each time step. Greenberg defined context as "*a dynamic construct*" [3], which is in accordance with the dynamic nature of group interactions in meetings.

We present a set of context ontology for context representation in our framework. The overall context should be composed of all the information related to human, physical and information environment. As is illustrated in Fig. 3, context of multi-party human interactions in the meeting room can be divided into three categories: (1) "*environment related context*", which represents the situation of physical environment; (2) "*information system context*" that is related to the situation of information system; (3) "*human-environment relation context*", which represents the relative relationship between human and physical environment; (4) "*human related context*" that denotes historical and current information about human.

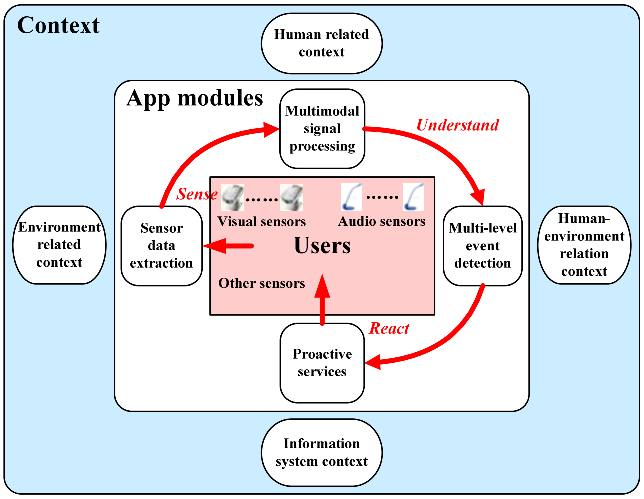


Figure 3. Context ontology for multi-party human interactions in meeting room

In this paper context awareness mainly refers to the "*human related context*". In [11] group interactive patterns were categorized based on different temporal scales and group sizes. In our current work the group size is fixed and we only try to categorize group interactions according to different temporal scales. As is shown in Fig. 4, "*human related context*" of group meetings can be divided into multiple levels according to various temporal scales and abstract semantic levels. For instance, a group meeting may contain three types of group situations "*presentation*", "*discussion*" and "*break*" at shorter time scales. A "*discussion*" stage may also contain three types of sub-situations "*two-member interaction*", "*all-member interaction*" and "*group voting*". In a sub-stage "*all-member interaction*", the discussion procedure might contain various speaking sections, such as "*A talking*", "*B talking*" etc. At the bottom level, multiple sub-sections such as "*B addressing to A*", "*B addressing to C*" etc may appear in a "*B talking*" section.

Currently we do not take group interest levels or individual mental states such as attitudes and expressions into account, i.e. we only consider human physical actions and interactions. Thus in our existed framework, current "*human related context*" can be expressed as a concatenation of context hierarchy, for instance, during meetings there might be a situation like this: "*during discussion, all members are interacting with each other, currently B is talking, and addressing to C*", which is an instance of Fig. 4. In the future work, we will extend the "*human related context*" from human physical actions and interactions to human mental states.

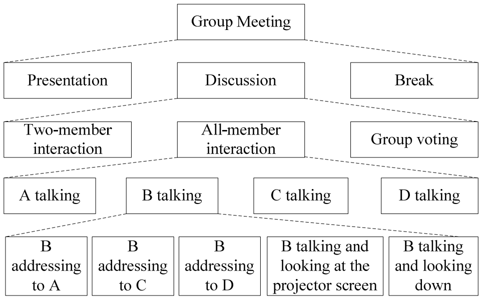


Figure 4. Hierarchy of human related context in small group meetings according to temporal scales and abstract levels

2.3 Context Awareness in Assistive Meeting Environment

As is introduced above, context awareness is the key of implicit HCI and assistive environment. In other words, context-aware computing serves as a bridge between the physical space and the cyberspace. Computer system understands users' actions and intentions in the physical space, and provides appropriate services from the cyberspace to the users automatically.

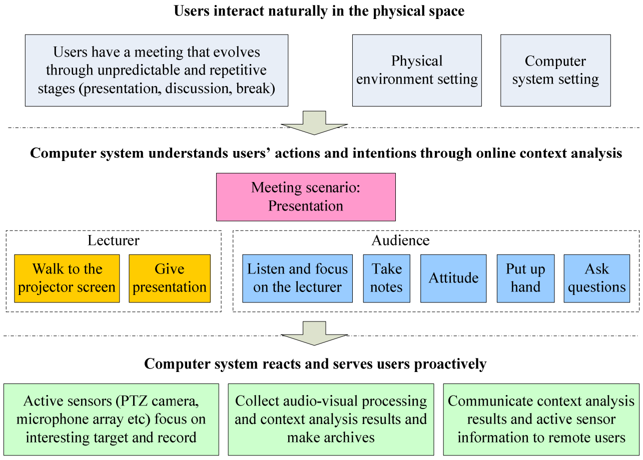

Users interact naturally in the physical space
Users have a meeting that evolves through unpredictable and repetitive stages (presentation, discussion, break)
Physical environment setting
Computer system setting
Computer system understands users' actions and intentions through online context analysis
Meeting scenario: Presentation
Lecturer
Walk to the projector screen
Give presentation
Audience
Listen and focus on the lecturer
Take notes
Attitude
Put up hand
Ask questions
Computer system reacts and serves users proactively
Active sensors (PTZ camera, microphone array etc) focus on interesting target and record
Collect audio-visual processing and context analysis results and make archives
Communicate context analysis results and active sensor information to remote users

Figure 5. Context awareness in assistive meeting environment

Fig. 5 illustrates the basic idea of context awareness in our research of group meeting analysis and assistive services. Users interact with each other naturally in the physical space, and the meeting procedure evolves through repetitive stages of "*presentation*", "*discussion*" and "*break*", which is not a predefined sequence. Computer system understands users' actions and the overall group situations through online context analysis. Based on the analysis results, computer system reacts and serves users proactively. The attentive services include: (1) active sensors (Pan-Tilt-Zoom camera and microphone arrays) focus on the most important object and record information; (2) make archives automatically online based on the multimodal processing and context analysis results; (3) communicate current processed results and active sensor information to remote meeting participants. From the flowchart we can see that context reasoning is the foundation of context-aware computing.

3. DYNAMIC CONTEXT MODEL

In order to achieve human-centered computing objectives in assistive meeting environment, we present a Dynamic Context Model as our solution strategy for context-aware computing.

The structure of our Dynamic Context Model is illustrated in Fig. 6. Dynamic Context Model is comprised of several major parts: context representation, context sensing engine and context guiding engine. Context is represented based on the context ontology introduced in Section 2. As is shown in Fig. 6, "*environment related context*", "*information system context*", "*human-environment relation context*" and "*human related context*" are recorded and maintained in our framework. Within the context sensing engine, multi-level events are detected based on multiple cues, which include lower level events and multimodal features extracted by various signal processing modules. Furthermore, detected events will serve as "*human related context*" and provide context-aware guidance through the context guiding engine. The top-down context guidance includes: (1) achieving selectivity in multimodal signal processing, (2) control of active sensors, (3) guidance for context sensing procedures.

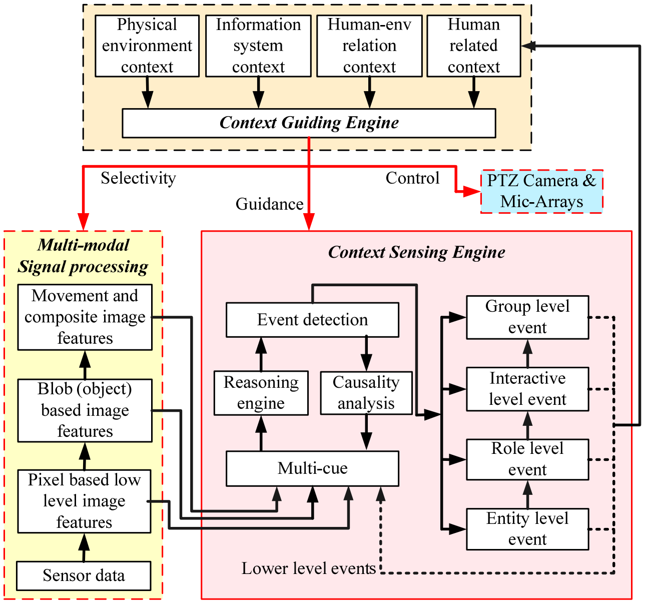


Figure 6. Dynamic Context Model

3.1 Multimodal Signal Processing

To infer the current context, the first significant step is to process the audio-visual sensor data so as to extract human locations, speaker directions, and even locations of body parts. Video and audio streams are extracted from multiple cameras and microphone arrays installed in the room and taken as the input of our multimodal signal processing system. Motion blobs and skin blobs are detected as the first two steps, based on which head and body blobs are initialized. These detected head and body blobs are tracked by multiple Particle Filter based tracking modules. Compared to other related work, our approach has the advantage of tracking adjustment and re-initialization by taking human records and current context into account. Detected or tracked human objects are maintained in the human records, which can be used for the re-initialization or adjustment of tracking modules. Furthermore, current context can be applied as the guidance of other refined visual processing modules. For some specific situations and roles, visual processing modules such as hand blob detection, head pose estimation etc will be initialized, so that the computer system can understand what the human is expressing with his hands or where he is looking at.

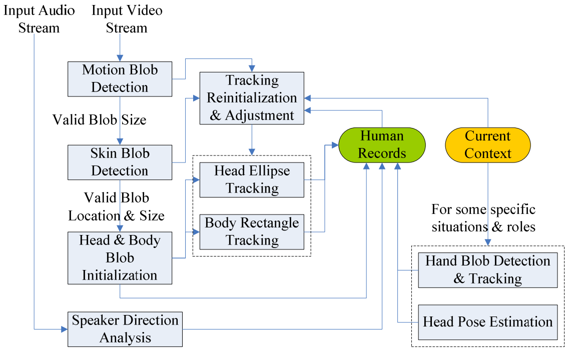

Input Audio Stream
Input Video Stream
Motion Blob Detection
Tracking Reinitialization & Adjustment
Valid Blob Size
Skin Blob Detection
Human Records
Current Context
Head Ellipse Tracking
Valid Blob Location & Size
Body Rectangle Tracking
For some specific situations & roles
Head & Body Blob Initialization
Hand Blob Detection & Tracking
Speaker Direction Analysis
Head Pose Estimation

Figure 7. Multimodal signal processing flowchart

The entire flowchart of our multimodal signal processing mechanism is illustrated in Fig. 7. Our work with context aware visual processing mechanism has been introduced in [12] and [13].

3.2 Context Sensing Engine

Context sensing engine holds the key to the operation of our framework. At any time step, computer system needs to perform online reasoning of current context. In this paper we mainly focus on "*human related context*", i.e. our objective is to understand what happens within the interaction groups. As shown in Fig. 4, "*human related context*" of group meetings can be divided into multiple levels according to various temporal scales and abstract levels.

Context and events are closely related to each other, since events are defined under specific context, and context is inferred through detecting events. In [5] events are defined to represent changes in situation that can be used to trigger system actions. In this paper we extend this concept and define two types of events: "*switching event*" that results in situation changes, and "*characteristic event*" that characterizes current situation and does not trigger situation changes. Furthermore, in our work events that might happen during multi-party interactions are defined to be comprised of four abstract levels: group, interactive, role and entity level, which correspond to the hierarchy of "*human related context*".

The relationship between "*human related context*" and events is illustrated in Fig. 8. Four context levels are expressed according to various temporal scales. At context level 1, given the group meeting scenario, three types of group events can be detected, which further play the role of context at context level 2. Interactive events are detected given the context level 1 and 2. At the bottom level, detected entity events play the role of context at context level 4 and are applied to guide selectivity and fusion of multimodal features and audio-visual processing modules. As a whole, the event hierarchy detected at current time step is to play the role of context in event detection at next time step.

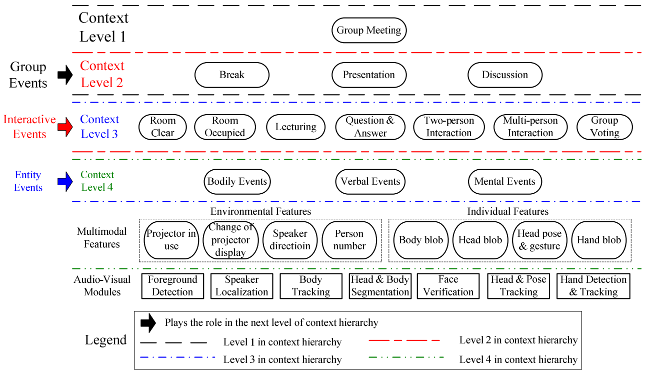


Figure 8. Event hierarchy for probabilistic modeling

Current context is sensed by detecting and integrating events at all hierarchical levels. Lower level events and multimodal features are used as multiple cues for the inference of higher level events, while higher level events play the role of context in detecting lower level events. According to this characteristic, we present a probabilistic model to detect multi-level events online, which is a multi-level event driven Dynamic Bayesian Network. Each level of events in Fig. 8 are modeled as a hierarchical level in the probabilistic model, and the multimodal features extracted from various audio-visual sensors are integrated as the observation for the inference of the multi-level state nodes in the DBN model. More details about our work on the probabilistic modeling of event hierarchy were introduced in [14].

3.3 Context Guiding Engine

Current context is not only essential for the understanding of human actions and intentions, but also can be used to guide the computer system to focus on the most important objects and tasks, which can reduce the work load of the system with limited resources. According to this idea, context guiding engine is proposed to perform top-down guidance for the computer system according to current context.

Multi-level events are detected in a probabilistic model, therefore current context is also uncertain. Context guiding engine performs its functions in two different ways. Firstly, the detected current and historical events are adopted as the probabilistic context and used in the DBN model for event detection. Secondly, we choose the combination of multi-level events with the maximum probability as fixed context, based on which a rule-based approach is employed to: (1) generate control commands for active sensors to focus on those significant objects in the scenario, and (2) select reasonable combination of signal processing modules to generate multimodal features.

Several context guiding examples are listed in Table 1, which includes the guidance for both the active sensor control and the signal processing module selection. However we did not integrate active sensor control into our system, which is to be extended in our future work.

Table 1. Context guiding examples

Current context	Active sensor control	Signal processing module selection
Break scenario, one person A in the room	No control commands for the active sensors	Track the head and body of person A, detect if another person arises
Presentation scenario, A gives lecture, B raises hand and asks question	PTZ camera turns to B, microphone arrays enhance the B direction	Detect the head pose of B, detect and track the hand motion of B
Discussion scenario, C and D interact with each other, C talks	PTZ camera turns to C, microphone arrays enhance the C direction	Detect the head pose of C, detect and track the hand motion of C, detect the head gesture of D

4. IMPLEMENTATION FRAMEWORK

We set up a smart meeting room to serve as our research platform. The meeting room setting has been introduced in Section 2.

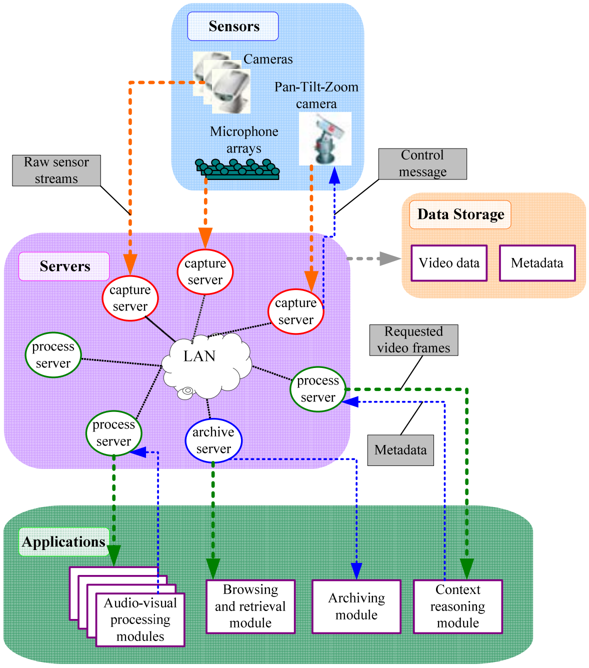
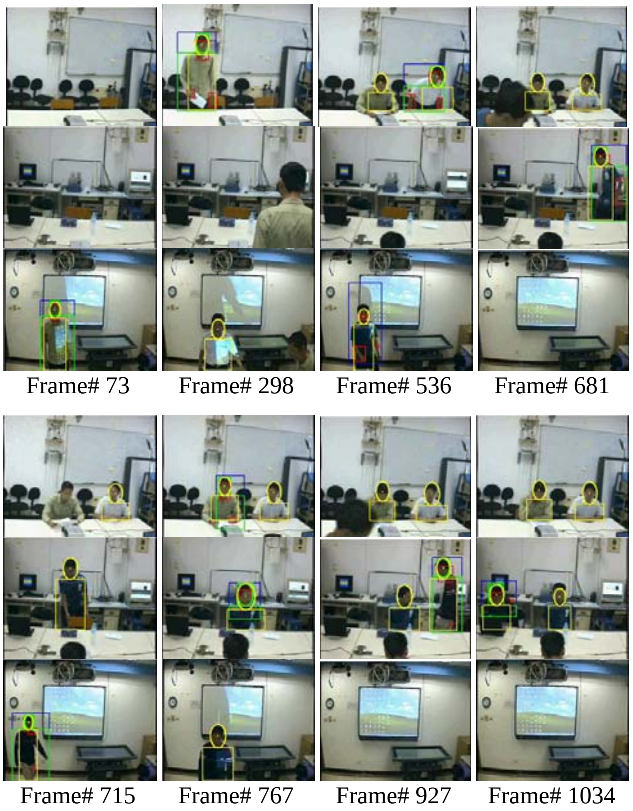


Figure 9. Implemented system framework

Based on the multiple audio-visual sensors distributed in our smart meeting room, a flexible multi-server platform is proposed as the software infrastructure to support distributed multimedia information processing, which include such functions as sensor data capture, transmission, analysis, storage and retrieval [15]. The distributed framework makes our system more efficient and robust. Multiple processing modules are distributed on various servers and can operate simultaneously. Furthermore, failure of one module will not result in the crash of other modules and the entire system.

Fig. 9 gives an overview of the system framework in our implementation. The software architecture is divided into several domains: server domain, sensor domain, application domain, and data storage domain. In the server domain, multiple servers with various functions are interconnected, which form the basis of the system. Regarding the sensor domain, fixed cameras, PTZ camera and microphone arrays are managed and connected to various servers. Within the application domain, application modules are maintained independently, which include signal processing modules, context reasoning module, retrieval and browsing module, and archiving module. Various modules communicate with each other between different servers through LAN, and transmitted data can be divided into three types: raw sensor streams, metadata, and control command. In the data storage domain, video data and metadata is archived in MPEG7 format.

5. EXPERIMENTAL RESULTS

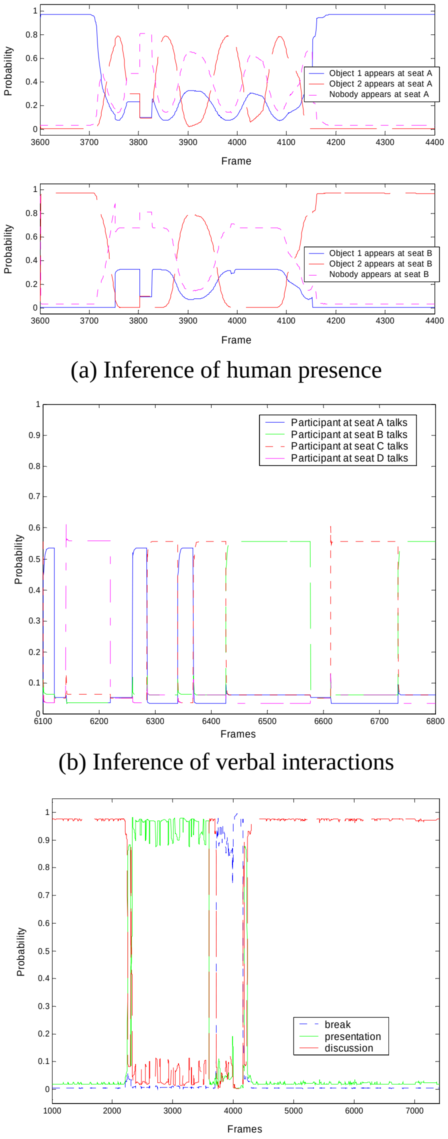


Frame# 73 Frame# 298 Frame# 536 Frame# 681

Frame# 715 Frame# 767 Frame# 927 Frame# 1034

Figure 10. Visual processing results

Some preliminary experimental results are introduced in this section. Multiple four-member group meeting sequences have been recorded in our meeting room, based on which we have made preliminary experiments and tested the framework proposed in this paper. Video sequences are extracted by three fixed cameras at distinct perspectives, and speech signals are recorded with three linear microphone arrays on the meeting table. Video sequences and speech signals are synchronized.



(a) Inference of human presence

(b) Inference of verbal interactions

(c) Inference of group situations

Figure 11. Context-aware group meeting analysis results

Audio processing module is relatively simple, which infers speaker directions based on the audio intensity and audio source angles recorded by the microphone arrays. However, visual processing is more challenging due to the complexity of the scenario and the limitations of each visual processing module. Visual processing results with the three fixed cameras are given in Fig. 10, in which each row corresponds to a camera. Our framework integrates multiple detection and tracking modules effectively and can handle tracking adjustment and re-initialization problems. As is shown in Fig. 10, the left participant in the first camera is occluded by another participant at Frame 681 and the tracking module lose the object at Frame 715, however our algorithm successfully finds the object and restarts the tracking process again at Frame 767.

Audio-visual processing generates a group of multimodal features, which can be used for the inference of current context. In our framework, current context is inferred through detecting multi-level events. Fig. 11 gives the inference results of some events. Fig. 11(a) shows the human presence detection results at two seats around the table in a period of "*break*" scenario. A "*discussion*" scenario is given in Fig. 11(b), from which we can see the turn taking sequence and four participants talk interactively with each other. The overall group situations are inferred at the top level of the DBN model at the results are given in Fig. 11(c), which segment the group meeting into multiple phases of "*presentation*", "*discussion*" and "*break*". All the events are detected in an online mode, i.e. the DBN model runs in a frame-by-frame style and gives inference results of multi-level events at each time step.

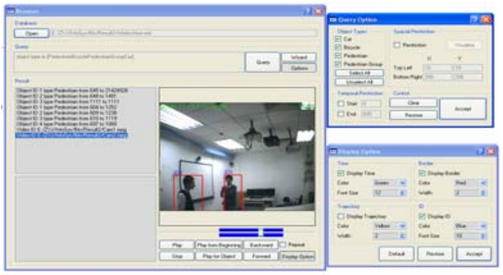


Figure 12. Retrieval UI

The archived information includes the recorded audio and video streams, signal processing results, and context reasoning results. All the processing results are archived in MPEG7 format, which can be used for future retrievals. Fig. 12 shows the user interface for information retrieval in our system.

6. CONCLUSION

Context has to be considered while analyzing human actions and intentions in the domain of implicit human computer interaction and pervasive computing. Sensing contextual information in multi-party human interactions such as small group meetings is especially challenging due to the dynamic and hierarchical nature of "*human related context*". This paper presents a novel Dynamic Context Model as our solution strategy for context-aware computing in group meeting analysis and assistive services. A distributed multimedia processing framework is implemented to realize our concepts. The proposed system has been demonstrated to be effective with preliminary experiments in the meeting room.

Future work includes improving and extending the probabilistic event detection model, and integrating active sensors such as PTZ camera into our framework.

7. ACKNOWLEDGMENTS

The work described in this paper is supported by CNSF grant 60673189 and CNSF grant 60433030.

8. REFERENCES

[1] Pantic, M., Pentland, A., Nijholt, A., and Huang, T. Human Computing and Machine Understanding of Human Behavior: A Survey. In *Proceedings of 8th International Conference on Multimodal Interfaces*, 2006, 239-248.

[2] Dourish, P., and Bellotti, V. Awareness and Coordination in Shared Workspaces. In *Proceedings of ACM Conference on Computer Supported Cooperative Work,* Toronto, Ontario, ACM Press, 1992.

[3] Greenberg, S. Context as a Dynamic Construct. *Journal of Human-Computer Interaction*, 16, 2001, 257-268.

[4] Baldauf, M., Dustdar, S., and Rosenberg, F. A Survey on Context-Aware Systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2006.

[5] Tan, J. G., Zhang, D., Wang, X., and Cheng, H. S. Enhancing Semantic Spaces with Event-Driven Context Interpretation. In *Proceedings of PERVASIVE*, LNCS 3468, 2005, 80-97.

[6] Crowley, J. L. Context Driven Observation of Human Activity. In *Proceedings of European Symposium on Ambient Intelligence*, 2003.

[7] Trivedi, M. M., Huang, K. S., and Mikic, I. Dynamic Context Capture and Distributed Video Arrays for Intelligent Spaces. *IEEE Trans. on Systems, Man, and Cybernetics—PART A: Systems and Humans*, 35, 1, 2005, 145-163.

[8] McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., and Zhang, D. Automatic Analysis of Multimodal Group Actions in Meetings. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27, 3, 2005, 305-317.

[9] Zhang, D., Gatica-Perez, D., Bengio, S., and McCowan, I. Modeling individual and group actions in meetings with layered HMMs. *IEEE Trans. on Multimedia*, 8, 3, 2006, 509-520.

[10] Dielmann, A., and Renals, S. Automatic Meeting Segmentation Using Dynamic Bayesian Networks. *IEEE Transactions on Multimedia*, 9, 1, 2007, 25-36.

[11] Gatica-Perez, D. Analyzing Group Interactions in Conversations: a Review. In *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Heidelberg, Germany, Sep. 3-6, 2006, 41-46.

[12] Dai, P., Tao, L., and Xu, G. Dynamic Context Driven Human Detection and Tracking in Meeting Scenarios. In *Proceedings of 2nd International Conference on Computer Vision Theory and Applications*, Volume Special Session, 2007, 31-38.

[13] Dai, P., Tao, L., Zhang, X., Dong, L., and Xu, G. An Adaptive Vision System toward Implicit Human Computer Interaction. In *Proceedings of 12th International Conference on Human-Computer Interaction*, Universal Access in HCI, Part II, 2007, 792-801.

[14] Dai, P., Di, H., Dong, L., Tao, L., and Xu, G. Group Interaction Analysis in Dynamic Context. In *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics (TSMCB)*, Vol. 38, No. 1, Feb. 2008, pp. 275 - 282.

[15] Wang, Y., Tao, L., Liu, Q., Zhao, Y., and Xu, G. A flexible multi-server platform for distributed video information processing. In *Proceedings of 5th International Conference on Computer Vision Systems*, Germany, 2007.