

# Exploiting SenseCam for Helping the Blind in Business Negotiations

Shuaib Karim, Amin Andjomshoaa, A. Min Tjoa

Vienna University of Technology  
{skarim, andjo, amin}@ifs.tuwien.ac.at

**Abstract.** During business meetings, blind persons are not able to see the meaningful movements, and facial gestures of the participants. The formal meeting minutes and / or participants' conversation during the meeting normally lack this important feedback in order to determine who is in favor and who is against their proposed suggestions. This is crucial in business negotiations, where one has to convince people and do lobbying for winning the business case in upcoming meetings. Today the devices already exist for instantly and seamlessly capturing the snapshots everywhere. The proposition suggests data capture using a similar device called SenseCam<sup>1</sup>, and then making these snapshots accessible for the benefit of the visually impaired users.

## 1 Introduction

Meeting room is a place of great deal of multi-modal knowledge exchange and decision making. The archiving of meeting proceedings is considered to be very useful for the immediate benefit of the participants. However, exploring the associations between the meeting room constituents vastly increases its benefits. By meeting room constituents we mean the participants, and objects within the meeting room. Multiple forums and vocabularies, and Multi-media information integration are marked as two of the technical research challenges by [1]. For meeting recognition, the important clues are marked as *What is being discussed, Who is in action, To whom one is talking to, When the meeting is taking place, and Where* [2].

The above questions can be significantly answered by representing the associations of personal information in ontologies. Our system explores and manages the associations existing within user's lifetime information using semantic web technology. There are various types of associations. *Firstly, there exist structural associations*, i.e., each information item has an inherent association with its structural metadata. For example, an email is associated with header fields such as Subject, From, To, Received / Sent Date, etc. *Secondly, associations could be asserted using manual annotations*. For example, a contact information item x can be manually associated with a project proposal document y based upon collaboration done by x on y. *Thirdly, using sophisticated techniques of textual and multimedia content analyses* further

---

<sup>1</sup> [http://research.microsoft.com/sendev/project\\_sensecam.aspx](http://research.microsoft.com/sendev/project_sensecam.aspx) (March 5, 2006)

associations in the concerned ontology are possible. The examples are the possible association of an email message containing the word "ICCHP" in body with some web page having title "ICCHP 2006", association of information items existing in the same time slice (editing a project proposal document & browsing the web for related information in parallel), association of information items generated from the same location or related with same location (a picture taken & a telephone call made from the same place). Also, significant benefit is achievable by finding the active / passive project participants based upon the presence / absence in related project meetings which can be calculated once the pictures are annotated for relative movements / gestures of the participants.

The next step is to present these associations in an accessible way for people with special needs. Generally for this purpose, accessibility is taken into account at the presentation layer according to the Web Content Accessibility Guidelines mentioned in WAI<sup>2</sup>. The guidelines as their name suggests are primarily focused on the content of web pages which is mostly textual HTML based information. Style sheets are used effectively to make the content and presentation separate and thus independent of each other. The dynamic contents are managed using client and server side scripts. Any non-textual objects in the page are made accessible by giving alternate text or long description in textual form by the web page authors. The scenario in the case of our implemented prototype personal information management system SemanticLIFE [3] is different though. Because the data feed is not only limited to web pages. The primary contents in our system are the user information items such as emails, documents of various types, chat sessions, telephone logs, process state data, browsed web history, contacts, calendar data and snapshots to be captured by SenseCam or some other similar device. When the associations are visualized at the interface the appropriate accessibility guidelines can be applied using the above mentioned techniques.

*The point of our interest is the need of applying accessibility criteria at the level of information contents and not just at the presentation level for facilitating the exploration of associations for people with special needs.* Moreover, the benefit is enhanced by making associations through manual annotations by applying it on captured snapshots.

In the next section related work is discussed including an overview of the SenseCam. Then some sample scenarios are described which we intend to fulfill, followed by proposed approach and workflow for managing associations.

## **2. Related Work**

The importance of the domain is highlighted by the ongoing projects related with meeting room recognition technologies such as Meeting Recorder Project by [4], Automatic Meeting Recognition Project by [1], Meeting Browser and Computers in the Human Interaction Loop by [5], Augmented Multiparty Interaction [6], and Interactive Multimodal Information Management [7], just to name a few.

---

<sup>2</sup> Web Accessibility Initiative of the W3C <http://www.w3.org/WAI/> (March 5, 2006)

Multimedia techniques have been used to track the meeting with participant's ID using color appearance, face id, and speaker id [2]. The corpus based framework in [8] describes how the meetings are modeled in layers, and how the annotation could be used to enhance significantly in meeting recognition. The work done by [9] in this area is highly related with our work. They have used image processing techniques to accomplish the task of monitoring and tracking the user activities. Instead we are focusing on linking the user's information items with each other based upon metadata to achieve the same goal. This will also help in resolving the synchronization issue between different data sources, a problem mentioned by [9]. The dialog act labeling guide [10] describes with particular emphasis the structure of audio dialog and the annotation system used for this purpose.

Our approach is somehow different because it is concerned with usage of annotations for exploiting the gestures and movements of participants instead of audio dialog. It is possible to identify and then annotate the meeting room constituents by using ontologies for interconnected information items. We are optimistic that this component will supplement the capture and recognition of meeting room knowledge exchange significantly.

There are various image annotation tools for the semantic web such as flick2rdf<sup>3</sup>, PhotoStuff<sup>4</sup>, SWAD<sup>5</sup>, M-OntoMat-Annotizer<sup>6</sup>. Some of these convert the annotations into owl file like flick2rdf. But the supported annotations which are mostly user comments are not sufficient for our purpose of capturing the metadata. The recent use case document by W3C [11] describes the issues and challenges in carrying out manual, semi-automatic, and completely automatic annotation of images. Our approach is in-line with their vision as we are starting with semi-automatic annotations and keeping our system architecture flexible to integrate the efforts done by image processing community in future.

For applying our strategy the snapshots of meeting must first be captured in sufficient detail. SenseCam is one such device capable of doing this. It is already under use by Microsoft in their research project MylifeBits [12] [13] with very positive results highlighting new challenges in managing personal information in various domains. A brief overview of the device is as follows:

#### **Overview of SenseCam.**

Designed by Lyndsay Williams, MSR Cambridge Lab., SenseCam is a badge size wearable camera that can automatically record your whole day's activities with a capacity of thousands of VGA images per day. The sensitive sensors detect any small change in environment & trigger a snapshot being taken. The sensors are configurable through xml configuration file. It is possible to get the location & the time stamp for each picture. Currently the sensors detect a change in location, light & temperature. For example, change due to another person coming into the room, opening / closing of door, changing posture like sitting down, standing up, running etc. Also,

---

<sup>3</sup> <http://purl.org/net/kanzaki/flickr2rdf> (March 5, 2006)

<sup>4</sup> <http://www.mindswap.org/2003/PhotoStuff/> (March 5, 2006)

<sup>5</sup> <http://swordfish.rdfweb.org/discovery/2004/03/w3photo/annotate.html> (March 5, 2006)

<sup>6</sup> <http://www.acemedia.org/aceMedia/results/software/m-ontomat-annotizer.html> (March 5, 2006)

the snapshot is triggered upon hand gesture for providing user control. The sensor data is recorded per second. It has accelerometers for motion detection & image stabilization. Its version 2 also has GPS and continuous audio recording capability. Another interesting feature is its capability to detect the presence of other SenseCams in the vicinity. All the data captured is stored in database tables in Sqlserver. The captured data is accessible either using available API or by directly accessing the database tables.

Currently, the doctors in Cambridge are doing clinical trials for getting its benefits for patients suffering from memory problems for memory recall with the help of all the pictures taken during the day. However, we intend to use it for blind by exploiting the annotation mechanism in our prototype system [3].

### 3 Sample Scenarios

**Blind person sitting in a business meeting.** The possible movements by the participants are leaving or entering the room, sitting down, standing up, whispering with someone while leaning, relaxing on the chair, sitting alert, hand gestures by the participants etc. These are some of the meaningful movements along with facial gestures to estimate the mood of the participants during discussion.

- The blind person wears the SenseCam during the meeting. So the pictures of the whole proceeding are taken automatically.
- Pictures are uploaded in SemanticLIFE repository as our file upload data feed.
- Retrieval of day's pictures from the repository and identification of participants either manually by the caregiver or automatically [14] using multimedia analysis plugins with possible help from ontologies.
- Annotation of pictures by the caregiver based upon the gestures / movements of the constituents (participants & other objects in the vicinity). Initially, it is assumed that the caregiver is already informed about the identities of the constituting objects. Later, the identities should be matched against similarity using ontologies.
- Following the annotations, enrichment of the associations would be made. For example, the gestures by a participant would update his / her contact profile for a particular project meeting. This will give useful information to the blind user about this participant in future meetings.

The system is usable in domains other than project meetings & also for a more diverse range of special needs [15]. The following scenario is useful for blind persons as well as for mobility impaired users. However, this is not the main focus for this paper.

**Blind person visiting the city.** In this case, the captured data is much broader. The pictures of our interest can also be those of stationary objects like restaurants, information counters, monuments, facilities coming on way to the destination.

- Blind person equipped with SenseCam is traveling from one station to another while the pictures are being taken continuously. The pictures would most probably

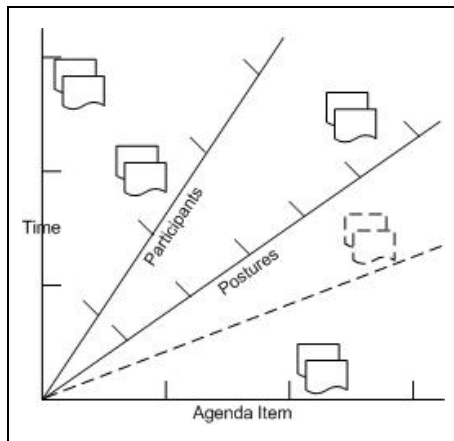
include the snapshots of the facilities like restrooms, lifts, ramps about which the blind person is not aware of on his / her initial visit. However, once captured the subsequent travel planning and travel itself could become more accessible.

- Similar to the above scenario, the pictures would be annotated by his / her caregiver, and associations would be made using ontology.
- Before making the subsequent visit, the blind user can consult his / her route planning. Based upon the associations made in the previous step, the system could present to him the information about availability of accessible facilities en route.

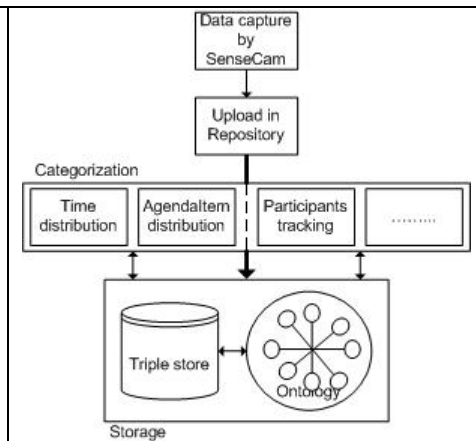
*Since, pictures taken by SenseCam are at the rate of about 1 picture / sec. Therefore, for a proceeding of 1 hour, there could be Approx. 3600 pictures which is a big number. The annotation is to be made as user friendly as possible so that the caregiver is able to annotate the whole day's activities easily.*

#### 4 Proposed Approach

The big amount of pictures is categorized into distinct parts based upon meaningful criteria such as given in Fig.1. The parts thus holding a relatively smaller amount of photos are annotated which is a relatively manageable task. The criteria to categorize the parts are devised in such a way so as not to lose any meaningful information about the whole meeting proceedings, yet being able to retrieve everything based upon a group of criteria, like a data cube [16]. The primary criterion is time distribution. The secondary criterion is to change the part boundaries based upon the participants' tracking (their identification and relative postures). Another criterion is to associate the issues or agenda items discussed during the meeting with the parts. There can be more criteria in future, based upon evolving semantics and needs of the users.



**Fig. 1.** Information distribution over multiple axes



**Fig. 2.** Workflow of annotation subsystem

There also exists some trivial project related ontologies (see for example UMBC project<sup>7</sup>, and doap<sup>8</sup>), however these are not sufficient for our requirements as such, but can be extended. A *Project Meeting Ontology* is proposed which will be used to specify the project & meeting related concepts and the binding between them (see Fig. 3).

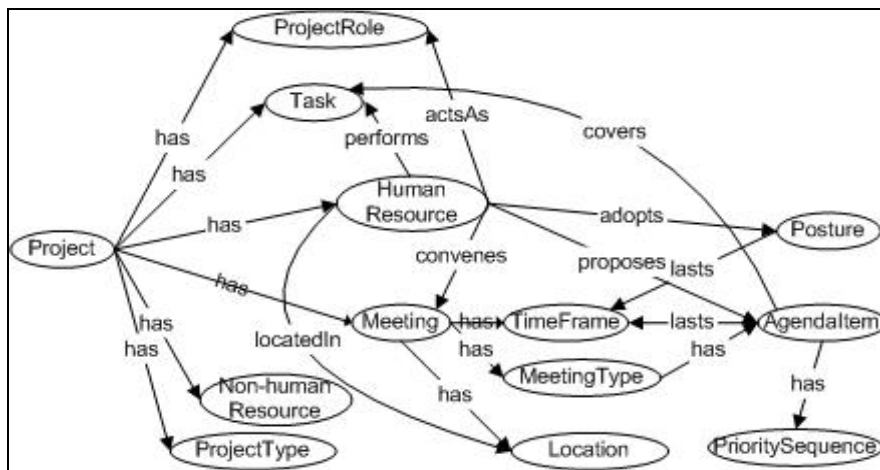


Fig. 3. A sample Project Meeting Ontology

The ontology is in its initial state. However it depicts the essential required concepts to fulfill our scenario. “MeetingTypes” can be highly formal, moderately formal, informal, research brainstorming, regular departmental meeting, coordination meeting etc. Meetings are carried out in context of some project(s). Each “Project” has its own dynamics. A “ProjectType” may range from software development, construction, tourism, entertainment etc. Each “ProjectType” has specific “ProjectRoles” such as manager, analyst, developer, tester, integrator etc. depending upon the project responsibilities. There exist typical “Tasks” associated for each type of role to carry out the responsibilities. The tasks are discussed in the meetings under specific “AgendaItems” lasting some “TimeFrame”. Normally, the meeting agenda is typical for specific type of project and specific type of meeting. Meeting is convened by someone (normally a “HumanResource”), and the meeting agenda items are also supposed to be proposed by someone. The meeting takes place at a certain “Location”, the person also has a location while attending the meeting (inside the meeting room or at remote location), the meeting has a “TimeFrame”, & so do the agenda items which are normally prioritized (have “PrioritySequence”) by someone before hand. Each “Agenda Item” lasts for some duration, & the participants assume “Postures” lasting some timeframe throughout the meeting. The postures can be; locally present, re-

<sup>7</sup> <http://ebiquity.umbc.edu/ontology/project.owl> (March 5, 2006)

<sup>8</sup> <http://usefulinc.com/doap> (March 5, 2006)

motely present, absent, standing, sitting, talking on telephone, working on laptop, apparently sleeping etc.

#### **4.1 Distribution according to time**

A long meeting can be heuristically broken down into  $n$  minute's duration of  $m$  parts each. For this purpose the timestamp available in EXIF header of the pictures is used. This process is carried out automatically by the analysis plugin in our system based upon user predefined values in the configuration file.

#### **4.2 Distribution according to participants' tracking**

The tracking of participants is very important because the issues discussed / decided in the meeting can be meaningfully related with it. The tracking can be divided into layers. Also, distribution can be made based upon identification of the participants irrespective of their postures. For example, distribution of pictures based upon names / identifications of the persons. Initially it is sufficient to tag the exit / entry of a participant which can be identified by following a manual protocol, or using RFIDs. Later, some of the major postures (as described in section 5.1) lasting for the duration of sections can be identified and annotated.

#### **4.3 Distribution according to issues**

The nature of the issues or AgendaItems is specific from project to project. A lot of issues are of generic nature and would thus be common in many projects. There is a small chance that the issues are unique for a common project type and meeting type.

### **5 Workflow for Managing Associations**

The three types of associations (structural, manual, dynamic) mentioned in Section 1, are taken care of for the stored pictures in order to make these accessible for blind. The information items are stored as RDF triples in SemanticLIFE repository against a trivial ontology. This base ontology specifies the header fields for each information item as its properties. The user interface plays an important role for convenient interaction and annotation of the pictures. The different types of information items from this triple store are displayed differently, say in different colors and on different graph. According to user studies the whole day is generally divided into 4 to 15 different activities out of which the meetings are generally 2 to 5 for the users under study. The activities with their times and durations may already be described in the calendar. With the help of range sliders user is able to select the slices of timeline

during which a meeting has taken place. Sometimes, the activities planned may overlap with activities carried out. In this case, user can handle it by adjusting and confirming the slider positions on the interface. In the selected time slice user is able to filter out the information items other than the pictures (taken by SenseCam). Each selected time slice may correspond to Subject of the appointment in calendar. If it is not specified over there, then upon a right click of the mouse, user is given option to put the items of the selected time slice into a new named collection, say SemanticLIFEMeeting of type “Meeting” which can possibly be connected to ontology of Meeting specified elsewhere on the web. Generally, the meeting participants for this meeting can be retrieved from the calendar, if specified over there. Otherwise, upon right clicking the SemanticLIFEMeeting, the user could enter the meeting participants.

### **5.1 Structural enhancement of information item**

By visualizing at the retrieved pictures the components of the meeting room constituents are described by the caregiver. The Exif headers for the pictures contain data about Camera-specific properties (make, model, sensing method, lens size etc.) and data about Image-specific properties (creation date, image resolution, height, width etc.). The structure can be enriched to take into account the constituents, e.g., describe who is present in the picture, and their postures (sitting, standing, sitting down, standing up, talking, writing, reading, working on laptop, attending a telephone call etc.). We agree that this is a laborious and time taking task for the caregiver to describe many pictures. With the help of intuitive user interfaces, and exploiting ontology of interconnected information items, the task can be made more suggestive and convenient for the user. Now when the information item is presented to the user this added constituents’ information along with existing Exif header information could be forwarded to assistive technology like screen reader for the blind. This feature alone would improve the learning experience of the user with impairments.

### **5.2 Manual associations with other information items**

It is also possible that some other activity like a telephonic conversation, chat session or web browsing was carried out during the meeting. These activities may or may not be related with the meeting under progress. Using manual annotations the individual pictures or collections of pictures can be associated with each other or any other information item in the repository or a concept. For example rating some named collections as highly useful, moderately useful and useless. This would later help in analyzing the usage of your time, and other project participants.

### **5.3 Dynamic associations**

Once the ontology (see Fig 3) instances are populated, then dynamic association of entities is possible. The information extraction techniques will be applied to the con-



tents of fed information in cases where linking of fields in item header is insufficient to firmly establish the associations. For this purpose concepts / key-terms are mapped with the ontologies. The links established through the ontology would make it possible to see the historical behavior of a person in specific type of meetings, historical analysis of specific agenda items with the duration of discussion etc.

## 6 Conclusions and Future Work

The automatic data capture devices like SenseCam do have the capability to capture much of the user's activities. However, it is still far from building an automatic diary for the user. An initial ontology of information items is developed and tested on our prototype system SemanticLIFE [3] for making associations. A proposal has been suggested to supplement the annotation of snapshots, and then building automatic associations using ontologies in the meeting domain. A projects meeting ontology is proposed. This would be of particular benefit for blind people and generally beneficial for people with diverse needs.

The work is still in progress. There are many privacy issues which we intend to investigate once the initial corpus is built.

## Acknowledgements

This work is supported by the financial support of Austrian National Bank (OeNB) for BIM (Business Information Management) research project, the ASEAN-EU University Network Initiative, and the Higher Education Commission of Pakistan.

## References

- [1] NIST Automatic Meeting Room Recognition Project, [http://www.nist.gov/speech/test\\_beds/mr\\_proj/index.html](http://www.nist.gov/speech/test_beds/mr_proj/index.html) (March 5, 2006)
- [2] Yang J., Zhu X., Gross R., Kominek J., Pan Y. and Waibel A., Multimodal People ID for a Multimedia Meeting Browse, Proceedings of the ACM Multimedia 1999, pp. 159-168.
- [3] Ahmed M., Hoang H.H., Karim M.S., Khusro S., Lanzenberger M., Latif K., Michlmayr E., Mustofa K., Nguyen H.T, Rauber A., Schatten A., Tho M.N. and Tjoa A.M., 'SemanticLIFE' - A Framework for Managing Information of A Human Lifetime, Proceedings of the International Conference on Information Integration, Web-Applications and Services, IIWAS'04, 27-29th Sep. 2004 Jakarta-Indonesia.
- [4] ICSI Meeting Recorder Project, <http://www.icsi.berkeley.edu/Speech/mr/> (March 5, 2006).
- [5] interACT Meeting Room Project, [http://penance.is.cs.cmu.edu/meeting\\_room/](http://penance.is.cs.cmu.edu/meeting_room/) (March 5, 2006).
- [6] Augmented Multi-Party Interaction, <http://www.amiproject.org/> (March 5, 2006).
- [7] NCCR Interactive Multimodal Information Management Project, <http://www.im2.ch/> (March 5, 2006).

- [8] Reidsma D., Rienks R.J. and Jovanovic N., Meeting modeling in the context of Multimodal research, Proceedings of the MLMI'04, LNCS 3361 (2005), pp. 22-35.
- [9] Wactlar H., Bharucha A., Stevens S., Hauptmann A. and Christel M., A System of Video Information Capture, Indexing and Retrieval for Interpreting Human Activity, IEEE International Symposium on Image and Signal Processing and Analysis (ISPA'03), Special Session on System Perspectives in Information Retrieval, Rome, Italy, September 18-20, 2003.
- [10] Elizabeth S., Raj D., Sonali B., Jeremy A. and Hannah C., The ICSI Meeting Recorder Dialog Act (MRDA) Corpus, Proceedings of SIGDIAL'04 (5th SIGdial Workshop on Discourse and Dialogue), 2004, pp. 97-100.
- [11] Image Annotation on the Semantic Web,  
[http://www.w3.org/2001/sw/BestPractices/MM/image\\_annotation.html#use\\_cases](http://www.w3.org/2001/sw/BestPractices/MM/image_annotation.html#use_cases)
- [12] Gemmell J., Bell G. and Lueder R., MyLifeBits: a personal database for everything, vol. 49, No. 1 Communications of the ACM, January 2006, pp. 88-95.
- [13] Gemmel J., Bell G., Lueder R., Drucker S., Wong C., MyLifeBits: Fulfilling the Memex Vision, Proceedings of ACM Multimedia '02, 2002, pp. 235-238.
- [14] Jeon J., Lavrenko V. and Manmatha R., Automatic image annotation and retrieval using cross-media relevance models, Proceedings of the 26th International ACM SIGIR Conference SIGIR 2003, ACM 2003, pp. 119-126.
- [15] Pühretmair F., It's Time to Make eTourism Accessible, Proceedings of the 9th International Conference on Computers Helping People with Special Needs ICCHP 2004, LNCS 3118 (2004), pp. 272-279.
- [16] Gray J., Chaudhuri S., Bosworth A., Layman A., Reichart D., Venkatrao M., Pellow F. and Pirahesh H., Data Cube: A Relational Aggregation Operator Generalizing Group-by, Cross-Tab and Sub Totals, Data Mining and Knowledge Discovery Journal, vol. 1, No. 1, 1997, pp. 29-53.