Chapter 9

User Study

While the evaluation of individual stages show that each stage has an accuracy between 60% and 98%, it is still essential to conduct a user study where the user retrieve their own experiences in a home environment. We conducted such a user study with a family who stayed in ubiquitous home during one of the real-life experiments. The coming sections of this chapter describe this user study and the results.

9.1 Objectives

The user study was designed with the following objectives:

- Identify requirements for experience retrieval in the ubiquitous home, using feedback from people who lived in the environment.
- Evaluate algorithms that have been implemented so far for video summarization and retrieval.
- Evaluate the usability of the system as a multimedia mining tool for use by residents of a household with a non-technical background.
- Identify directions for future work and improvement of the existing algorithms.

Since there was no previously designed evaluation experiment that fulfills the above objectives, we designed a detailed user study that is specific to our system. Since the system was being developed at the time the user study was conducted, we used a prototype system that did not have some of the functionality described in Chapters 6

and 7. However, this was not a problem or weakness as one of the objectives of the user study was to identify the user requirements, and the feedback received could be utilized for implementing a better system.

9.2 Participants

We selected a family who stayed in ubiquitous home for two weeks, during one of the real-life experiments. The family had three members; a married couple, and their 3-year old daughter. On the working days, the husband went to work and the wife stayed most of the time at home, taking care of the child. The family went out occasionally for meals, shopping and other family activities, but returned every night as usual for an ordinary household. There were a few guests, but they did not stay overnight. The subjects were paid for participation in the user study.

9.3 Procedure

This study consisted of two parts; a requirements analysis for home experience retrieval, and a hands-on session of the system we developed. Data captured during six hours on the 12th of April 2005 (the 4th day of the family's stay in ubiquitous home) were used for the study. This was equivalent to retrieving from 102 hours of video and 150 hours of audio data. The date of the user study was the 12th of October 2005, exactly 6 months later.

The experiment consisted of three sections. The first section was a requirement study, where the subjects answered a questionnaire to specify what they would expect from a system for retrieving experiences at home. This section of the experiment was

conducted before demonstrating the system, to ensure that the user requirements are not influenced by the functionality of the existing system.

In the second section, the subjects were given a demonstration on how to use the system. Only one example for each type of retrieval was shown. Thereafter, they were allowed to use the system themselves, submitting their own queries to retrieve their experiences. The authors were available in case the subjects needed advice, but were not involved in using the system. The subjects were asked to select video clips that they would like to keep, so that we can provide them in a DVD. This was done both as a factor of motivation and also to find out what kind of experiences generate interest in keeping a permanent record. After using the system, the subjects rated the usability of the system by answering a brief questionnaire based on the guidelines by Chin et al.[74].

In the third section, the subjects provided descriptive feedback about the system. The subjects were asked to suggest additional requirements to what they proposed previously, in case if there were any.

The user study took approximately 3 hours, and the subjects were paid for participation. The subjects provided their responses in separate answer sheets but used the system together. This helped to elicit more responses, rather than getting only those both subjects agreed upon. Since the child is only 3 years old, only the parents actively participated in the experiment. Other than for restarting the system due to an operating system crash, no assistance was needed from the authors.

9.4 Results

The questions and the responses for the requirement study are stated below. The number of subjects who provided each answer is stated in parentheses.

1. Suppose it is possible to retrieve any event that happened anytime during your stay in ubiquitous home. What are the things you would like to see from that stay?
    - Things that I did (2)
    - Things that the other family members did (2)
    - How my child was playing when she was alone (2)
    - Things that I have forgotten (2)
    - Things we did together (1)
    - A summary of what I did each day (1)

The subjects added the following after using the system.

- Recall what we did when my friend visited the house (1)
- See my own behavior and habits, e.g.: gait (1)
- See my child growing up over a long period of time (1)

2. Supposing it is possible to have the same facility at your home, and only your family has access to the data:

    (a) What would you like to use it for?

    - For taking care of my old mother; check whether she took medicine properly, or she ate too much sweets, etc. (1)
    - For finding lost objects, discover our own habits, find out how the child is behaving so that anything bad can be corrected (1)

(b) How would you like to record the data?

- Record everything from daily life (2)

(c) Which parts of the house will you record everything?

- Non-private areas of the house, like the living room, the kitchen, etc. (2)
- Child's room (1)

(d) Which parts of the house and times will you refrain from recording?

- Private places, such as bedrooms (2)

(e) Do you have any other preferences, such as times of day, about recording?

- Would like to see my child during day time and afternoon, when I am not at home(1)
- Want to record leisurely times playing with the child (1)
- Would like to record busy hours of the day for discovering things that were left behind (1)

For the usability assessment, the following were the responses from the two subjects on a seven-point response scale with 1 being the worst rating and 7 being the best.

- Learning to use the system – 6,6
- Ease of using the system – 4,5

- Overall impression – 5,6

The following are the questions in the section for feedback about the system, and the responses from the subjects.

1. How much did you remember from what you could see in the video and key frames?
   - There were many things that I did not remember. For example, that I worked that day (1)
   - I roughly remembered what happened on that day. But the memory was refreshed a lot after watching the video (1)

2. Was it possible to see interesting things that you did not see/know before?
   - Yes (2)
   - We could discover things like how our child woke up in the morning (1)
   - I was surprised to learn that I spend so much time with my child (1)

3. Out of what you saw, which parts of the video would you like to keep with you?
   - Video of the child (2)
   - Video of activities we did together, such as having meals (1)

4. State what you like about this software.
   - Automatic camera change (1)
   - Ability to see what happened when I was away (1)

5. State what you don't like about it.
    - It might reveal things that are not nice to know (1)
    - Too many video clips and key frames to look at (1)

6. For what kind of things will this software be useful to you?
    - Family diary, security
    - Taking care of family members
    - Record of our child's life
    - See myself objectively.

The responses to the requirements show that the system can already match most of the requirements the subjects had in their mind before using it. The subjects found the system easy to use, as suggested by the high rankings for the usability assessment. Descriptive feedback indicates that the subjects found the software useful and it helped them to discover a few things that they were not conscious about or did not know at all.

The subjects managed to recall what happened in the entire session and to retrieve video they wanted to watch, by using the system. They found two types of video more interesting, and watched them repeatedly. One type contained video of the child when she was alone: an example was the video clip created when the child woke up in the morning, found that she was alone in the living room, and ran for the mother. The other type corresponded to activities that they did together, such as taking meals and playing with the child. They requested copies of both these types of video clips. The subjects used key frames as an index to the original video, rather than viewing only the key

frames as a summary. They liked using the system, and it was somewhat difficult to get them to stop watching videos and answer the questionnaire.

9.5 Discussion

The behaviors of residents in the two types of experiments were significantly different. While the subjects in *students' experiments* were independent in their actions, the behavior of the family in the real-life experiment was in the form of a group. This affected the quality of the results, too. For student experiments, video clips and summaries resulting from handover and key frame extraction were mostly exclusive whereas those created during the real life experiment had a lot of overlap and redundancy, due to behavior as a group. For instance, when the child was following or walking by the side of a parent, the personalized video created for the child and the parent have near 100% overlap, which results in redundancy. Therefore, video retrieval for group behavior seems to be more important for a real-life situation. Furthermore, the accuracy of footstep segmentation decreases because of complex walking patterns created by a child walking with a parent. The accuracy is about 30% less than reported previously for students' experiments.

With only two persons actively taking part in the user study, the responses have little statistical value. However, their keen interest on using the system and positive feedback justifies the motivation and the current progress of this work. The responses also provide valuable insights to identify further requirements and possible improvements. Continuing further study with other families, as the system is being developed, will help the system to evolve into one that is very useful.

As the subjects indicated in their feedback, privacy of the residents should be protected by recording data only in the public locations of the house. Although this reduces the ability of the system to function as a memory assistant, it is an important measure as individual privacy is important even for the members of the same family. Furthermore, the system was helpful for the residents even with restrictions in locations. It can be suggested that one of the reasons for the success of the real-life experiments (in the sense that the residents enjoyed their stay and retrieval of their experiences) is that the residents were not confined to the house, and their privacy was protected even when they were in the house.

Chapter 10

Conclusion and Future Work

10.1 Conclusion

We have implemented video retrieval and summarization for a home with a large number of sensors, by analyzing both content and context data. While the inclusion of multiple sensory modalities and the human centered aspect of the system made the research fairly broad, it was possible to make a number of novel contributions.

Hierarchical clustering of floor sensor data followed by video handover enabled the creation of personalized video clips using a large number of cameras. It was possible to dub this video with reasonably good quality, using audio handover. An adaptive algorithm enabled retrieval of more than 80% of the key frames required for a complete summary of the video.

Silence elimination and false positive removal from audio data produced results with a high accuracy of 98%. The scaled template matching algorithm we propose is able to achieve generally accurate sound source localization despite the absence of microphone arrays or a beam-forming setup. The accuracy of audio classification using only time domain features is above 83%, suggesting that high accuracy of classification is possible at the expense of further analysis using features from multiple domains.

Basic image analysis facilitated detection of events that are useful in understanding the activities that take place inside the house. Action detection using multiple sensory modalities yielded an average accuracy of approximately 78%.

The user interface based on hierarchical media segmentation and Interactive retrieval facilitated effective retrieval with a small amount of manual data input using only a pointing device. Visualizations of different types of data at various levels of detail helped the user to retrieve required media.

The residents who evaluated the system found it useful, and enjoyed using it. They found the system easy to learn and usable. The requirements they identified and the feedback they provided were valuable in improving the system.

10.2 Future Work

The accuracy of the existing algorithm for footstep segmentation can be improved for more efficient retrieval of personalized video. Adding person recognition capability to this stage will enable the creation of *personal diaries* for residents.

Use of frequency and Cepstral domain features for similarity matching is a prospective approach for improving the performance of sound source localization. Applying Independent Component Analysis (ICA) before Audio classification to separate multiple simultaneous sound sources within the same region is an interesting research direction.

Further analysis of sensor data, especially image data, can be used for detection and recognition of higher-level actions and events, such as conversations, mealtimes, etc., thereby enhancing the functionality of the system. Face detection in retrieved images and video can provide additional information for searching within the data.

Novel techniques for user interaction and visualization of results can be designed, to achieve more effective and efficient retrieval. Continuous evaluation of the system

based on user studies and feedback is vital, as usability is one of the most important criteria for an effective retrieval system for the home.

References

[1] Pingali, G., Jain, R., "Electronic Chronicles: Empowering Individuals, Groups, and Organizations", Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Amsterdam, July 2005.

[2] Pentland, A. "Smart Rooms," Scientific American, pp.54-62, 1996.

[3] Harper, R., "Inside the Smart Home: Ideas, Possibilities, and Methods", Springer-Verlag UK, July 2003.

[4] Abowd, G. D., Bobick, I., Essa, I., Mynatt, E., Rogers, W. The Aware Home: Developing Technologies for Successful Aging. American Assoc. of Artificial Intelligence (AAAI) Conf, 2002.

[5] Noguchi, K.; Somwong, P.; Matsubara, T.; Nakauchi, Y., "Human intention detection and activity support system for ubiquitous autonomy," *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on* , vol.2, no.pp. 906- 911 vol.2, 16-20 July 2003.

[6] Brumitt, B., "Easy Living: Technologies for Intelligent Environments," Proc. of International Symposium on Handheld and Ubiquitous Computing, 2000.

[7] Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S. "Multi-Camera Multi-Person Tracking for Easy Living," Proc. of 3rd IEEE International Workshop on Visual Surveillance, pp.3-10, 2000.

[8] Lee, J., Ando, N., Hashimoto, H., "Design Policy for Intelligent Space," Proc. of IEEE SMC'99, 1999.

[9] Juels, A., "RFID security and privacy: a research survey," *Selected Areas in Communications, IEEE Journal on* , vol.24, no.2pp. 381- 394, Feb. 2006

[10] Fishkin, K. P., Wang, M., Borriello, G., "A ubiquitous system for medication monitoring," Intel Res., Seattle, Tech. Memo IRS-TR-03-011, Oct., 25 2004.

[11] Lee, S., Mase, K.,. "Activity and Location Recognition Using Wearable Sensors", Pervasive Computing, July-September, pp.10-18, September, 2002.

[12] Petrushin, V. A., Gang W., Omer S., Damian R., Gershman, V., "Multiple-Sensor Indoor Surveillance System," *crv*, p. 40, The 3rd Canadian Conference on Computer and Robot Vision (CRV'06), 2006.

[13] Lee, D., Hull, J. J., Erol, B., "Meeting Video Retrieval using Dynamic HMM Model Similarity", IEEE ICME Conference, 2005.

[14] Jaimes, A., Nagamine, T., Liu, J., Omura, K., Sebe, N., "Affective Meeting Video Analysis", IEEE International Conference on Multimedia and Expo (ICME'05), Amsterdam, The Netherlands, July 2005.

[15] Lijun T., Kender, J.R., "Semantic Indexing for Instructional Video Via Combination of Handwriting Recognition and Information Retrieval," *Multimedia*

and Expo, 2005. ICME 2005. IEEE International Conference on , vol., no.pp. 920- 923, 6-8 July 2005

[16] Bibiloni, A., & Galli, R. (2003) Content Based Retrieval Video System for Educational Purposes. *Proceedings of Eurographics Workshop on Multimedia "Multimedia on the Net, EGMM 1996,*

[17] Waibel, A., "CHIL – Computers in the Human Interaction Loop", In proceedings of LEARNTEC 2005.

[18] Cucchiara, R. 2005. Multimedia surveillance systems. In Proceedings of the Third ACM international Workshop on Video Surveillance &Amp; Sensor Networks (Hilton, Singapore, November 11 - 11, 2005). VSSN '05. ACM Press, New York, NY, 3-10.

[19] Gandhi, T., Trivedi, M. M., "Calibration of a reconfigurable array of omnidirectional cameras using a moving person, Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks, October 15-15, 2004, New York, NY, USA

[20] Lam, K., Chiu, K. H. C., Adaptive visual object surveillance with continuously moving panning camera, Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks, October 15-15, 2004, New York, NY, USA

[21] Valera, M., Velastin, S.A., Intelligent distributed surveillance systems: a review vision. In Image and Signal Processing, IEE Proceedings, volume 152, pages 192--204, April 2005.

[22] Y.A. Ivanov and A.F. Bobick, Recognition of Multi-agent Interactions in Video Surveillance," IEEE Proceedings of the International Conference on Computer Vision, Kerkyra, Greece, Vol. 1, pp. 169-176, September 20-27, 1999.

[23] J. Davis and A. Tyagi, "Minimal-Latency Human Action Recognition using Reliable-Inference", *Image and Vision Computing*, Vol 24, 2006, pp. 455-472.

[24] Collins, R.T., Lipton, A.J., Fujiyoshi, Kanade, T., "Algorithms for cooperative multisensor surveillance", In Proc. of the IEEE, volume 89, pp. 1456-1477, Oct. 2001.

[25] Brooks, R.R., Ramanathan, P., Sayeed, A.M., "Distributed target classification and tracking in sensor networks. In Proc. of the IEEE", volume 91, pp. 1163-1171, 2003.

[26] Arampatzis, Th.; Lygeros, J.; Manesis, S., "A Survey of Applications of Wireless Sensors and Wireless Sensor Networks," *Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterranean Conference on Control and Automation* , pp. 719- 724, 2005

[27] Khalifa, Y.; Okoene, E., "A distributed agent-based surveillance system," *Signals, Circuits and Systems, 2005. ISSCS 2005. International Symposium on*, vol.2, no.pp. 713- 716 Vol. 2, 14-15 July 2005

[28] Girgensohn, F., Shipman, S., Dunnigan, A., Turner, T., Wilcox, L., "Support for Effective Use of Multiple Video Streams in Security" , In Proceedings of the

Fourth ACM international Workshop on Video Surveillance & Sensor Networks , VSSN '06. ACM Press, New York, NY, pp.19-26.

[29] Wang, Y., Liu, Z., Huang, J., "Multimedia Content Analysis Using Both Audio and Visual Cues", IEEE Signal Processing Magazine, November 2000. 12-36

[30] Jaimes, A., Christel, M., Gilles, S., Sarukkai, R., Ma, W. Y. Multimedia Information Retrieval: What is it, and why isn't anyone using it? ACM MIR 2005.

[31] Cai, R., Lu, L., Hanjalic, A., "Unsupervised Content Discovery in Composite Audio", Proceedings of ACM Multimedia 2005, Singapore. (2005) Page 628-637

[32] Pye, D., Hollinghurst, N., Mills, T. and Wood, K., "Audio-visual Segmentation for Content-based Retrieval", The International Conference on Spoken Language Processing (ICSLP'98), Sydney, Australia, December 1998

[33] Sung I. Yong , Won S. Lee, Content-based retrieval of video data with flexibly managed attributes, Knowledge and Information Systems, v.4 n.4, p.507-519, October 2002

[34] Kim, S., Hwang, D. S., Kim, J., Seo, Y., "An Effective News Anchorperson Shot Detection Method Based on Adaptive Audio/Visual Model Generation", Proceedings of the Fourth International Conference in Image and Video Retrieval (CIVR 2005), Singapore. (2005) 276-285

[35] Brown, M.G., Foote, J.T., Jones, G.J.F., Sparck Jones, K., and Young, S.J., 1995, "Automatic Content-based Retrieval of Broadcast News," Proceedings of ACM Multimedia. San Francisco: ACM, pp. 35 - 43.

[36] Chien Yong Low , Qi Tian , Hongjiang Zhang, An automatic news video parsing, indexing and browsing system, Proceedings of the fourth ACM international conference on Multimedia, p.425-426, November 18-22, 1996

[37] Duan, L., Xu, M., Tian, Q., Xu, C., Jin, J.S. "A unified framework for semantic shot classification in sports video", IEEE Transactions on Multimedia, Volume: 7, Issue: 6, Dec. 2005. pp.1066-1083

[38] Tjondronegoro, Dian and Chen, Yi-Ping Phoebe and Pham, Binh (2002) *A Framework for Customizable Sport Video Management and Retrieval*, in Za¨ıane, Osmar and Simoff, Simeon and Djeraba, Chabane, Eds. *Mining Multimedia and Complex Data: KDD Workshop MDM/KDD 2002, PAKDD Workshop KDMCD 2002 Revised Papers*, pages pp. 248-265. Springer.

[39] Huayong, L. and Hui, Z. 2005. A Content-Based Broadcasted Sports Video Retrieval System Using Multiple Modalities: SportBR. In *Proceedings of the Fifth international Conference on Computer and information Technology* (September 21 - 23, 2005). CIT. IEEE Computer Society, Washington, DC, 652-656.

[40] Petkovic, M., Jonker, W., "Cobra: A Content-Based Video Retrieval System", Advances in Database Technology - EDBT 2002 : 8th International Conference on Extending Database Technology, Prague, Czech Republic, March 25-27, 2002. Proceedings, pp. 736-738.

[41] Davis, M. King, S. Good, N. From Context to Content: Leveraging Context to Infer Media Metadata. ACM Multimedia 2004, 188-195.

[42] Department of Sensory Media - Ubiquitous Sensor Room, http://www.mis.atr.jp/~megumu/IM_Web/MisIM-E.html, ATR Media Information Science Laboratories, Kyoto, Japan.

[43] Jaimes, A. Omura, K., Nagamine, T., Hirata, K. Memory Cues for Meeting Video Retrieval. CARPE 2004, 74-85.

[44] Mori, T., Noguchi, H., Takada, A., Sato, T. Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior. First International Workshop on Networked Sensing Systems (INSS2004), 40-43.

[45] Matsuoka, K., Fukushima, K. Understanding of Living Activity in a House for Real-time Life Support. SCIS & ISIS 2004, 1-6.

[46] Bush, V., :As We May Think", The Atlantic Monthly, 176(1), July 1945, pp.101-108.

[47] S. Mann, "Humanistic Intelligence: `WearComp' as a new framework and application for intelligent signal processing", Proceedings of the IEEE, Vol. 86, No. 11, November, 1998

[48] Jennifer Healey , Rosalind W. Picard, StartleCam: A Cybernetic Wearable Camera, Proceedings of the 2nd IEEE International Symposium on Wearable Computers, p.42, October 19-20, 1998

[49] Wood, K., Fleck, R. & Williams, L (2004) Playing with SenseCam. Proceedings of Playing with Sensors: Exploring the boundaries of sensing for playful ubiquitous computing (W3), UbiComp 2004

[50] Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C., "MyLifeBits: fulfilling the Memex Vision", in Proceedings of the tenth ACM international conference on Multimedia, ACM Press: Juan-les-Pins, France. pp. 235-238.

[51] Michael Beigl, Albert Krohn, Tobias Zimmer and Christian Decker: Typical Sensors needed in Ubiquitous and Pervasive Computing First International Workshop on Networked Sensing Systems (INSS) 2004, Tokyo, Japan, June 22-23. 2004, pp 153-158

[52] Aizawa, K., Kawasaki, S., Tancharoen, D., Yamasaki, T. Efficient Retrieval of Life Log based on Context and Content. ACM CARPE, 2004.

[53] Gemmell, J., Bell, G., Lueder, R., "MyLifeBits: a personal database for everything", Communications of the ACM, vol. 49, Issue 1 (Jan 2006), pp. 88-95.

[54] Ubiquitous Home: http://www.nict.go.jp/jt/a135/eng/research/ubiquitous_home.html, National Institute of Information and Communication Technology, Japan.

[55] Smeaton, A.F., McHugh, M., "Towards Event Detection in an Audio-Based Sensor Network", Proceedings of the Third ACM Workshop on Video Surveillance and Sensor Networks (VSSN'05), Singapore. (2005). pp.87-94

[56] Rui, Y., Florencio, D., "New direct approaches to robust sound source localization", IEEE International Conference on Multimedia & Expo (ICME 2003), Baltimore, USA. (2003)

[57] Chen, J. F., Shue, L., Sun, H. W., Phua, K. S. "An Adaptive Microphone Array with Local Acoustic Sensitivity", IEEE International Conference on Multimedia & Expo (ICME 2005), Amsterdam, The Netherlands. (2005)

[58] Hoshuyama, A. Sugiyama, A. Hirano, T., "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters", IEEE Trans. Signal Processing, vol. 47, no. 10, pp. 2677--2684, Oct. 1999.

[59] Vacher, M., Istrate, D., Besacier, L., Castelli, E., Serignat, J. F., "Smart audio sensor for telemedicine", Smart Objects Conference 2003, Grenoble, France. (2003)

[60] Bian, X., Abowd, G. D., Rehg, J. M., "Using Sound Source Localization to Monitor and Infer Activities in the Home", GVU Technical Report; GIT-GVU-04-20, Georgia Institute of Technology. (2004)

[61] Liu, M., Wan, C., "A Study on Content-Based Classification and Retrieval of Audio Database", Proceedings of the 2001 International Database Engineering and Applications Symposium, Grenoble, France. (2001) pp. 339-345

[62] Vazirgiannis, M., Sellis, T., "*Event and Action Representation and Composition for Multimedia Application Scenario Modelling*", Proceedings of the European Workshop on Interactive Distributed Multimedia Systems and Services (IDMS'96), March 1996.

[63] Snoek, G.M., Worring, M., "Multimedia Event-Based Video Indexing using Time Intervals", IEEE Transactions on Multimedia, 7(4) August 2005. pp.638-647,

[64] Kraft, F., Malkin, R., Schaaf, T., Waibel, A., "Temporal ICA for Classification of Acoustic Events in a Kitchen Environment," in INTERSPEECH, Lisbon, Portugal, 2005.

[65] de Silva, G.C., "Tracking and Indexing of Human Actions in Video Image Sequences", M. Eng. Thesis, Dept. of Electrical and Computer Engineering, the National University of Singapore, 2003.

[66] Wang, J. R. and Parameswaran, N. 2004. Survey of sports video analysis: research issues and applications. In *Proceedings of the Pan-Sydney Area Workshop on Visual information Processing*

[67] Piccardi, M., Hintz, T., He, S., Huang, M. L., Feng, D. D., Eds. ACM International Conference Proceeding Series, vol. 100. Australian Computer Society, Darlinghurst, Australia, pp.87-90.

[68] Witten, I. H., Frank, E., (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[69] Viola, P., Jones, M., *Robust Real-time Object Detection*, International Journal of Computer Vision, 2002. pp.1026-1033

[70] Anavi, S., "Information Processing in a Ubiquitous Home Using Image Analysis", Masters Thesis, Signal Processing Institute, School of Engineering Swiss Federal Institute of Technology, Lausanne, February 2006.

[71] Zhao, R., and Grosky, W. I., "Negotiating The Semantic Gap: From Feature Maps to Semantic Landscapes", Pattern Recognition, Volume 35, Number 3 (March 2002), pp. 51-58.

[72] Sommerville, I., "Software Engineering: 5th Edition", Addison-Wesley, 1993.

[73] Pingali, G., "eChronicles: Scope, Opportunities, and Challenges", Proceedings of IEEE eChronicle Workshop, 2006, pp.1-6.

[74] Chin, J. P., Diehl V. A., Norman, K. L. Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. Proceedings of ACM CHI'88 Conference on Human Factors in Computing Systems, 1988 pp.213-218.

Appendix A

Material used for Evaluation of Key Frame Extraction

Annexed starting from the following page are the instruction sheet, handouts and answer sheets used for the experiment for evaluating the algorithms for key frame extraction (Section 5.6). The key frames sets and the answer sheets for only one video clip are included, as an example.

Instruction Sheet

Overview

The objective of this experiment is to evaluate different methods of extracting *key frames* from a video clip. Key frames, in our context, are frames that are selected such that they present a **summary** of the events took place in the video clip. The motivation for creating a key frame set is saving time taken to view an entire video clip to find out what happened.

Please follow these instructions while taking part in the experiment. If there is anything that is not clear enough, please ask.

Instructions

- Before you start, please fill in your name and start time of the experiment, in the spaces provided in the answer sheet.
- You have to repeat the following steps on four video clips (sequences). You can take a break after each sequence.

1. Use the *Image Sequence Viewer* to watch the sequence. This is the original video, which is not summarized yet. Try using the buttons for playing, pausing and moving back and forth along the sequence. The durations of the sequences are:
    - Sequence 1 - 36 seconds
    - Sequence 2 - 1 minute
    - Sequence 3 - 3 minutes
    - Sequence 4 - 5 minutes
2. Select key frames for the sequence so that they represent a summary of the behavior of the person shown in the video clip. There is no restriction on the number of frames you can choose. There is no time limit for selecting key frames. Write the timestamps of the frames, in the sheet provided.

    Notes:

    - It is not necessary to write down the date component of the timestamp
        - Example: **2004/09/03 13:2:45:13** can be written down as **13:2:45:13**
    - It is not necessary to write the timestamps in any particular order. Feel free to revise and add more key frames. You may strike out the entries you made (Example: ~~13:12:24:5~~) if you feel that they are not necessary.

3. Now, observe each set of printed key frames and fill in the corresponding column in the table in Section A. There are seven sets of key frames for each sequence. You can go back to view the sequence as many times as necessary.
4. Proceed to Section C only after completing Section B.

- At the end of the experiment, please record the time in the space provided.

Experiment for Evaluation of Key Frame Extraction

Name : _____

Date : _____

Start time : _____

End time : _____

Test Sequence 1

Please use the numbers below the images to specify key frames.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

Sequence 1

Section A

Key frames selected for this sequence

Timestamp

Timestamp

Section B

Evaluation of individual key frame sets

	Key frame set						
	A	B	C	D	E	F	G
1. Number of key frames as compared to the duration of the sequence (a) Too few (b) Fine (c) Too many							
2. Percentage of redundant frames (a) None (b) Less than 25% (c) 25%-50% (d) More than 50%							
3. Number of important frames missed (a) None (b) 1 to 5 (c) 6 to 10 (d) More than 10							

Section C

2. Out of the 7 frame sets, which one do you think best key frame set for this sequence?

3. For the selected frame set:

   (a) Why do you find it better than other sequences?

   (b) What are the ways that it can be improved?

4. Additional comments for this sequence and key frame sets, if any:

Sequence 1

Frame set A

2004/9/3 13:41:2:28

2004/9/3 13:41:4:28

2004/9/3 13:41:5:28

2004/9/3 13:41:15:11

2004/9/3 13:41:29:17

2004/9/3 13:41:30:23

2004/9/3 13:41:33:23

Sequence 1

Frame set B

2004/9/3 13:41:1:22

2004/9/3 13:41:3:10

2004/9/3 13:41:5:4

2004/9/3 13:41:25:17

2004/9/3 13:41:30:23

2004/9/3 13:41:31:17

Sequence 1

Frame set C

2004/9/3 13:41:2:28

2004/9/3 13:41:4:28

2004/9/3 13:41:5:28

2004/9/3 13:41:31:29

2004/9/3 13:41:32:29

2004/9/3 13:41:38:29

Sequence 1

Frame set D

2004/9/3 13:41:1:16

2004/9/3 13:41:3:16

2004/9/3 13:41:5:16

2004/9/3 13:41:20:17

2004/9/3 13:41:30:17

2004/9/3 13:41:31:17

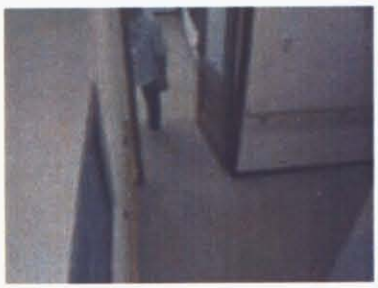
Sequence 1

Frame set E

2004/9/3 13:41:1:16

2004/9/3 13:41:3:16

2004/9/3 13:41:5:16



2004/9/3 13:41:30:17

2004/9/3 13:41:31:17

Sequence 1

Frame set F

2004/9/3 13:41:2:28

2004/9/3 13:41:17:17

2004/9/3 13:41:32:11

2004/9/3 13:41:38:29

Sequence 1

Frame set G

2004/9/3 13:41:2:28

2004/9/3 13:41:32:11

2004/9/3 13:41:38:29

Appendix B

Simplified Mathematical Model for Sound Source Localization

This appendix attempts to justify the scaled template matching using the energy distribution templates, using a simplified approach.

Let us assume that the sound energy received by the microphones within a given region of a closed environment is a linear combination of sound energy generated in each region, and no noise is present. The sound energy received within region m, r_m can be stated as

$$r_m = a_{m1}s_1 + a_{m2}s_2 + \ldots + s_m \ldots + a_{(n-1)}s_{(n-1)} + a_{mn}s_n$$

where n is the number of sources in the closed environment and S_i $(i = 1..n)$ is the sound energy released by the i^{th} sound source.

The ubiquitous home is partitioned into 7 regions where sounds can be generated. Sounds are captured by 25 microphones in 6 of those regions. For simplicity, we assume that we have only one *receiver* in each of these regions. The equations for received sound energy by the set of receivers can now be written in matrix form as

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} \\ a_{21} & 1 & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} \\ a_{31} & a_{32} & 1 & a_{34} & a_{35} & a_{36} & a_{37} \\ a_{41} & a_{42} & a_{43} & 1 & a_{45} & a_{46} & a_{47} \\ a_{51} & a_{52} & a_{53} & a_{54} & 1 & a_{56} & a_{57} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & 1 & a_{67} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{bmatrix}$$

$\Rightarrow \mathbf{r} = \mathbf{As}$

The matrix **A** represents the proportion of sound energy that propagates from one region to the other, in a normalized form. If matrix **A** can be obtained, the problem of sound source localization can be reduced to solving this matrix equation for **s** given **r** and **A**.

The energy distribution templates for the regions of ubiquitous home represent the propagation of sound energy to individual microphones of each region. Hence, each of the energy distribution templates serve as a detailed representation of each column of **A**.

However, it should be noted that the system cannot be solved by matrix inversion, since the number of elements in **r** is smaller than that of **s**. The scaled energy distribution template matching technique is equivalent to solving the set of linear equations for **s**.

Appendix C

Material used for the User Study

The annexed answer sheet was provided to the participants of the user study described in Chapter 9.

Answer Sheet

Name: .....

Part A: Requirements study

1. Suppose it is possible to retrieve any event that happened anytime during your stay in ubiquitous home. What are the things you would like to see from that stay? Please check any number of items as you like, and add your own choices.

   - ❑ Things that I did
   - ❑ Things that the other family members did
   - ❑ Things we did together
   - ❑ How my child was playing when she was alone
   - ❑ Things that I have forgotten
   - ❑ A summary of what I did each day
   - ❑ .....

     .....
   - ❑ .....

     .....
   - ❑ .....

     .....
   - ❑ .....

     .....

2. Supposing it is possible to have the same facility at your home, and only your family has access to the data:

(a) What would you like to use it for?

(b) How would you like to record the data?

- ❑ Record everything from daily life so that I can find out anything if I forget
- ❑ Record only the special events such as (please specify):

(c) Which parts of the house will you record everything?

- ❑ Everywhere
- ❑ The following places (please specify):

.....

.....

(d) Which parts of the house will you refrain from recoding?

.....

.....

(e) Do you have any other preferences, such as times of day, about recording?

.....

.....

Part B: Usability Ratings

Please answer the following questions after the demo and the hands-on session. Circle the number corresponding to your ranking for each criterion

Learning to use the system

1 Difficult	2	3	4	5	6	7 Easy

Overall impression

1 Difficult to use	2	3	4	5	6	7 Easy to use

1 Useless	2	3	4	5	6	7 Very useful

Part C: Descriptive Feedback

Please be descriptive as and when necessary when answering the following questions:

1. How much did you remember from what you could see in the video and key frames?

2. Was it possible to see interesting things that you did not see/know before?

3. Out of what you saw, which parts of video would you like to keep with you?

4. State what you like about this software.

5. State what you don' t like about it.

6. For what kind of things will this software be useful to you?