An Interactive Multimedia Diary for the Home

Gamhewage C. de Silva, Toshihiko Yamasaki, and Kiyoharu Aizawa
University of Tokyo

A system for retrieval and summarization of multimedia data from a home-like environment continuously captures video and audio sequences as the inhabitants are moving inside the house. An interactive user interface based on hierarchical media segmentation incorporates user memory and intelligence data. In the long term, this system can provide valuable information for studies related to housing design, evaluating human behavior, and so on.

With advances in technology, affordable methods for recording human experiences have become available, and the amount of multimedia content that individuals record has greatly increased over the past few decades. Combining this trend with the rapid growth in automated multimedia retrieval research makes investigating continuous capture and automated retrieval of personal experiences a timely endeavor.

Automated capture and retrieval of experiences taking place at home is interesting for several reasons. The home offers an environment where a variety of memorable events and experiences take place. Sometimes these events, such as a child's first footsteps, occur without an opportunity for manual capture. Sometimes, people do not want to remove themselves from the experience to shoot photos or video. Captured media can serve as a "memory assistant" to help recall things that were forgotten. In the long term, a corpus of interactions and experiences recorded at home can provide valuable information for studies related to the design of better housing, evaluating human behavior, and so on.

CHALLENGES

Capture and retrieval of multimedia in a home environment is a challenging task for several reasons. Even the simplest and the smallest houses are partitioned into rooms or defined areas, making it necessary to have several cameras and microphones to provide complete coverage. Continuous recording of data from these devices results in large amounts of data. The level of privacy inside the house is also a concern, depending on both the location and the persons involved.

The most difficult problem, however, arises during retrieval and summarization of the captured data. The content is much less structured compared to that from other environments. Retrieval queries can entail very different levels of complexity, and the results can encompass various levels of granularity. For example, possible queries might include the following:

- When did I get up on the first of this month?
- What time were we watching soccer last weekend?
- Who left the lights on in the study room last night?
- How did the strawberry jam I bought last week disappear in four days?

Unlike meeting or sports videos that are bounded by the time of the corresponding events, continuously captured data from a home environment is unbounded. This continuous capture increases the complexity of retrieval.

Given these issues and the state of the art of content-processing algorithms, multimedia retrieval for ubiquitous environments based solely on content analysis is neither

 Published by the IEEE Computer Society 0018-9162/07/$25.00

Multimedia Retrieval from Ubiquitous Environments

Ubiquitous environments are equipped with a large number of sensors of different types, enabling acquisition of data from events that take place within the area.

Current research on smart and ubiquitous environments can be divided into two major categories. One aims at providing services to the people in the environment by detecting and recognizing their actions. Examples include the Aware Home[1] and other Smart Home projects.[2] The other research category aims at storing and retrieving media at different levels ranging from photos to experiences. Recent developments in storage technologies that make it possible to record large amounts of data have facilitated this type of research. Some projects such as CHIL[3] attempt to combine both types of research by supporting real-time user interaction and using retrieval for long-term support.

Most of the existing research on multimedia retrieval deals with a previously edited single video stream with specific content.[4] For such data, the common approach is content analysis, using domain knowledge where applicable. However, using context data can greatly improve performance.[5] Researchers have used supplementary context information such as location, motion, and time to successfully index and retrieve life-log video captured by a wearable camera.[6]

Recent research on multimedia retrieval for ubiquitous environments includes development of the Ubiquitous Sensor Room,[7] an environment that captures data from both wearable and ubiquitous sensors to retrieve video diaries related to experiences of each person in a room. The Sensing Room[8] is a ubiquitous sensing environment equipped with cameras, floor sensors, and RFID sensors for long-term analysis of daily human behavior. This research segments video and sensor data into 10-minute intervals and uses a hidden Markov model to recognize the activity in the room during each segment. In an attempt to understand and support daily activity in a house, Katsunori Matsuoka and K. Fukushima[9] used a single camera installed in each room and sensors attached to the floor, furniture, and household appliances.

Some projects emphasize both human-human and human-computer interaction for efficient and accurate capture and retrieval. Masahi Takahashi and colleagues[10] proposed a layered framework for representing human interactions using machine readable indices. Alejandro Jaimes and coworkers[11] used graphical representations and interactive visualization of important memory cues for video retrieval from a ubiquitous environment. ■

References

1. G.D. Abowd et al., "The Aware Home: Developing Technologies for Successful Aging," *Proc. Am. Assoc. Artificial Intelligence Conf.*, AAAI Press, 2002, pp. 1-7.
2. R. Harper, *Inside the Smart Home: Ideas, Possibilities, and Methods*, Springer-Verlag, 2003.
3. A. Waibel, "CHIL–Computers in the Human Interaction Loop," *Proc. Learntec 2005*, Universität Karlsruhe (TH), 2005, pp. 2-9.
4. A. Jaimes et al., "Multimedia Information Retrieval: What Is It, and Why Isn't Anyone Using It?" *Proc. ACM MIR 2005,* ACM Press, 2005, pp. 3-8.
5. M. Davis, S. King, and N. Good, "From Context to Content: Leveraging Context to Infer Media Metadata," *Proc. ACM Multimedia 2004,* ACM Press, 2004, pp. 188-195.
6. K. Aizawa et al., "Efficient Retrieval of Life Log Based on Context and Content," *Proc. ACM CARPE*, ACM Press, 2004, pp. 22-31.
7. Department of Sensory Media, "Ubiquitous Sensor Room," ATR Media Information Science Laboratories, Kyoto, Japan; www.mis.atr.jp/past/im.html.
8. T. Mori et al., "Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior," *Proc. 1st Int'l Workshop Networked Sensing Systems* (INSS 2004), IEEE Press, 2004, pp. 40-43.
9. K. Matsuoka and K. Fukushima, "Understanding of Living Activity in a House for Real-Time Life Support," *Proc. SCIS & ISIS*, Tokyo Polytechnic Univ., 2004, pp. 1-6.
10. M. Takahashi et al., "A Layered Interpretation of Human Interaction Captured by Ubiquitous Sensors," *Proc. ACM CARPE 2003*, ACM Press, 2004, pp. 32-38.
11. A. Jaimes et al., "Memory Cues for Meeting Video Retrieval," *Proc ACM CARPE 2004*, ACM Press, 2005, pp. 74-85.

efficient nor accurate. Therefore, it is desirable to use supplementary data from other sensors for easier retrieval. Since ubiquitous environments are built with infrastructure to support cameras and microphones, adding sensors to acquire supplementary data is relatively easy.

UBIQUITOUS HOME AND MULTIMEDIA DIARY

Our work on multimedia capture and retrieval focuses on the development of algorithms for person tracking, key frame extraction, media handover, sound-source localization, lighting change detection, and the design and implementation of user interface interaction strategies that help to navigate huge amounts of multimedia data. The "Multimedia Retrieval from Ubiquitous Environments" sidebar describes related work in this area.

Our studies were conducted at the National Institute of Information and Communications Technology's Ubiquitous Home[1] in the Keihanna Human Info-Communication

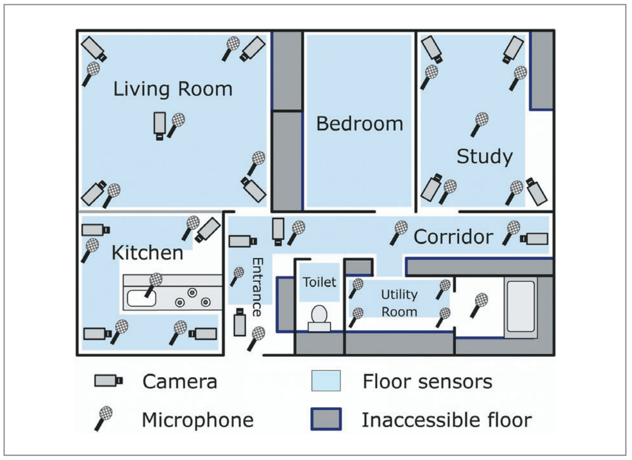

Figure 1. Ubiquitous Home sensor layout. The diagram shows the orientation of each camera. The microphones in the corners of the rooms have unidirectional responses toward the center of the room, whereas the other microphones are omnidirectional.

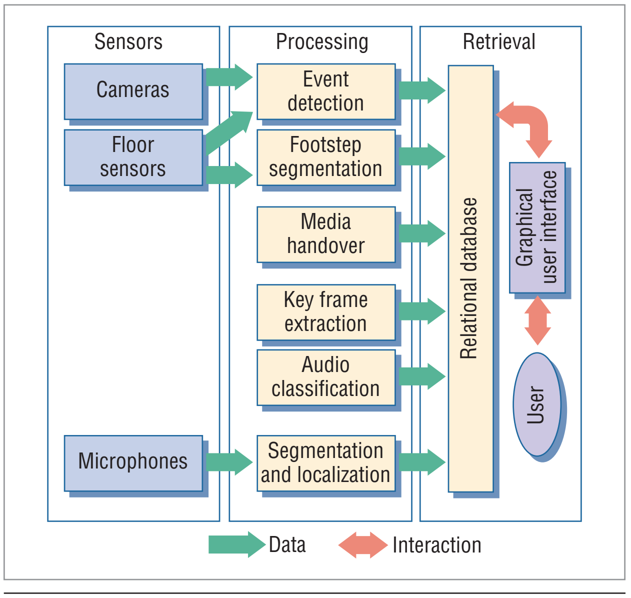

Figure 2. System overview. The sensor data is processed both together and independently, and the results are stored in a relational database. These results, accessible through the user interface, create the multimedia diary's index.

Research Center, Kyoto, Japan. Simulating a two-bedroom house, this environment is equipped with 17 cameras and 25 microphones for continuous video and audio acquisition. Pressure-based floor sensors are mounted in the areas shown in light blue in Figure 1.

The stationary cameras capture images at a rate of five frames per second and store them in JPEG file format. Audio is sampled at 44.1 kHz from each microphone and recorded in MP3 file format. The floor sensors are point-based pressure sensors spaced 180 mm apart in a rectangular grid. The sensors interface to a hardware controller that samples the pressure on each sensor at 6 Hz. A pair of binary state transitions occurs when weight is placed on and removed from a sensor. Each pair is combined to form sensor activation, thereby providing a way to represent a footstep using floor sensor data. The system records the sensor coordinates, starting time stamp, and the duration for each sensor activation.

Our Ubiquitous Home environment continuously captures this data to create a Multimedia Home Diary from which residents can use simple, interactive queries to retrieve stored data capturing their experience in the home.

DATA ANALYSIS

Our system, shown in Figure 2, analyzes data from floor sensors, microphones, and cameras for recognition of different types of actions and events and then writes the results to a central relational database, which users access through a graphical user interface by submitting interactive queries.

Floor sensor data retrieval

An agglomerative hierarchical clustering algorithm[2] segments the sensor activations into footstep sequences. As Figure 3a shows, the algorithm can segment footsteps of multiple persons walking in the house at the same time. The grid corresponds to the floor sensors, and the lighter shade of gray indicates sensor activations that occur later.

For each footstep sequence detected by our footstep segmentation algorithm, a position-based video handover algorithm[2] creates a video clip that keeps the person in view by changing cameras as the person moves inside the house. Figure 3b shows how the algorithm selects the cameras for the footstep sequence in Figure 3a. A lighter shade of the same color shows each camera's viewable region. The path's color indicates the camera that has been selected. As Figure 3c shows, audio handover is performed in a similar fashion to ensure that the person is heard throughout the video clip.

The video sequence constructed using video handover is also sampled to extract key frames, providing a summary of the video clip. An adaptive spatiotemporal sampling algorithm based on the elapsed time, camera transitions, and the rate of footsteps creates a complete compact summary.[2]

Audio data retrieval

As Figure 1 shows, the microphones capture sounds associated with regions of the house. We start by eliminating audio corresponding to silence in each region. The system performs silence elimination by comparing the audio signal's root mean square power against a threshold value. The threshold value for each microphone is based on the probabilistic distribution of one hour of audio data for silence for that microphone. A region-based voting algorithm that combines the sound seg-

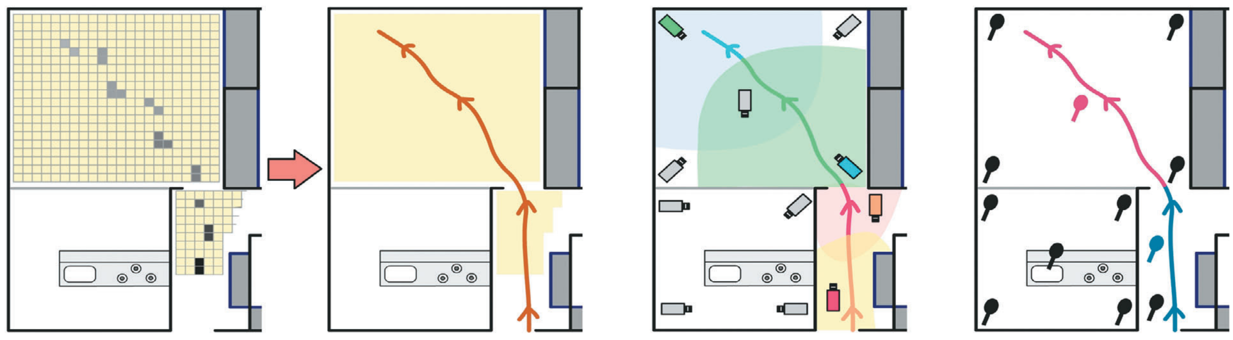

Figure 3. Key frame extraction. (a) The algorithm can segment footsteps of multiple persons walking in the house at the same time. (b) The algorithm selects the cameras for the footstep sequence. (c) Audio handover is performed in a similar fashion to ensure that the person is heard throughout the video clip.

ments from the microphones in the same region removes noise. The process results in a set of 1-second audio segments for situations in which a sound was heard at a given region.

The sounds contained in the resulting segments can be categorized into two types: *local sounds*—sounds generated in the same region as the microphone—and *overheard sounds*—sounds generated in a region other than the microphone's location. To prevent false retrievals, segments with only overheard sounds are removed before further processing. This *sound-source localization* task identifies the regions where one or more sound sources are present.

Lighting changes

If combined with the scene context, lighting changes can be used to identify significant events that take place in a house. Because they occur very quickly, they can be used to create short summaries of a day's events.

Lighting changes are relatively easy to detect as they are represented by sharp intensity-level changes in image sequences. However, the problem is finding a single threshold level for this change that yields accurate event detection. The light level in the images changes with the weather, curtains, time of day, and the automatic gain control in the cameras, which make the light level recover after the sharp changes.

To solve this problem, we assign a significance rank to each lighting change based on its sharpness. The user selects the rank and browses the events through an interactive interface, thereby intuitively reducing the search space.[3]

USER INTERACTION

The user retrieves video, audio, and key frames through a graphical user interface that accepts user queries and presents comprehensible results. However, there is a considerable difference between the user query and the semantic levels of the result. This semantic gap causes problems in most multimedia retrieval systems. If the gap is closed by making users submit lower-level queries, the system's usability decreases. On the other hand, trying to fill the gap using heuristics or simple assumptions will result in decreased accuracy.

We use a twofold approach to solve this problem. On the system side, hierarchical media segmentation provides better visualization of events. At the same time, interactive queries incorporate user intelligence into the query process.

Hierarchical media segmentation

The system continuously captures multimedia data, from which it retrieves segments corresponding to events as results. The system indexes original data from the Ubiquitous Home by source—camera ID, microphone ID, floor sensor coordinates—and time stamp. The results are indexed based on the time stamp, location, and event/action. Figure 4 shows the data segmentation after capture and analysis.

However, these indices still cannot facilitate efficient multimedia retrieval, as humans tend to remember events and query for them in a different way. For example, "Retrieve video showing the regions of the house people were in at 9:47 a.m." is a valid query for the system, but it is unlikely that a human user would enter such a query. A more likely query would be, "What was I doing after breakfast last Saturday?"

Humans tend to segment the time and experiences at home by days, sessions—for example, morning, late afternoon, evening—locations, and events. However, these are difficult to model using a computer-based system due to the complex semantic level and fuzzy boundaries of the segments. We propose a tradeoff that segments the media hierarchically by date, hour, location, and event. The user can browse the multimedia collection by viewing activity at each level of this hierarchy.

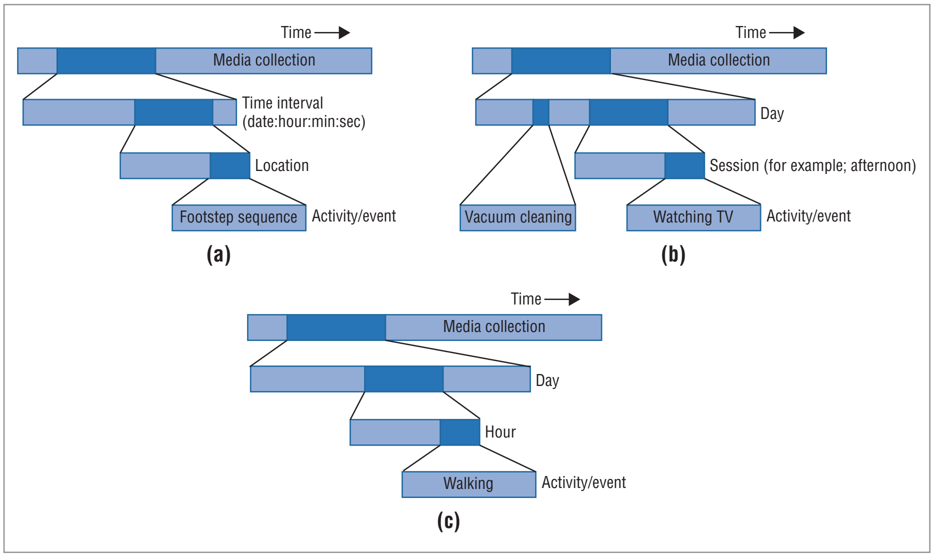

Figure 4. Hierarchical media segmentation. (a) Media segmentation based solely on indexes. (b) Media segmentation based solely on users' queries. (c) Proposed approach for media segmentation.

Interactive retrieval

In its current state and with the available sensor data, the system cannot perform some useful tasks such as person recognition. Further, the algorithms it implements are less than 100 percent accurate. However, incorporating user interactions into the system can greatly improve performance.

We use interactive retrieval to achieve this. The system breaks a query into several steps, and each step returns intermediate results so that the user can provide further input, navigating toward the desired results. For example, if the user wants to retrieve video showing what person A did during a given interval, the system provides a key frame from each video sequence created by segmenting footsteps. The user can look at the key frames and identify those showing person A, resulting in accurate retrieval at the expense of one additional step.

User interface design

We defined the user interface with a few key objectives. The first was to use only a pointing device (either a tablet monitor or a touch monitor) to facilitate faster input of queries and navigation through results compared to text-based interaction. The second was that the user needed minimal technical knowledge to understand the inputs and results—all user-adjustable parameters would be interpreted in a way the user could understand. For example, a slider control input labeled "Sampling rate gradient for key frames with range 0.0 to 1.0" has little meaning to the user. We rectified this by labeling the input as "Desired amount of key frames," with "few" and "many" labeling the slider's ends. The results were grouped according to different classifications such as the location and type of event to present them clearly and prevent information overload.

The third objective was to facilitate easy navigation within data without requiring the user to start over. Humans tend to search for relative queries such as "What happened next?" or "What happened in the other room during this time?" Our system facilitates such queries by dividing the results into a set of tabs and updating them according to the parameters submitted for the last query. While browsing the results, a user can navigate along the timeline outside the query boundaries, using button-based inputs such as "previous" and "next."

The user interface is designed to have more intuitive inputs. For example, the user can click on an image showing the home layout to select a room/region from which to retrieve events. Camera selection is facilitated in the same manner. The view from a camera is immediately available so that the user can select the right camera interactively. Instead of line or bar graphs, color levels are used to indicate each activity's intensity level because interpretation is faster and easier.

Event retrieval

Figure 5 shows an example of event retrieval using the system. Suppose the user wants to see what she did on a selected holiday. The user enters the date, and the system displays a summary of the day's activities along the timeline. The graphs are segmented in one-hour time intervals. Each graph's color level denotes the amount of corresponding activity. Colors are used instead of numbers to summarize each type of activity and event. The colors are scaled so that the visualizations show the results with maximum possible contrast. The activity patterns that Figure 5a shows are sufficient for the user to understand that she went to sleep around 1:00 a.m.

By clicking a selected one-hour block on a graph with the desired type of activity, the user can see a detailed summary for that hour. Figure 5b shows the retrieval of footstep sequences from midnight to 1:00 a.m. Showing the footsteps on the house plan, with the color changing from blue to red with time, allows the user to understand the movement without seeing the video, which expedites browsing.

Although the system does not have person recognition capability, since the initial video frame is displayed with each footstep sequence, the user can easily select the videos showing the right person. Figure 5c is a partial screenshot of a key frame, which can serve both as a summary of—and an index for—a video clip.

EVALUATION

A series of real-life experiments were conducted for data collection. We collected data from three families and controlled experiments done by three groups of students; however, only one family took part in the evaluation. The study participants included members of different ages. In each experiment, the participants lived in the Ubiquitous Home for a period of one to two weeks. They led their normal lives during this stay and were not restricted in terms of the amount of time they spent in the house. The participants went to work or school during weekdays, they cooked and had meals in the house, and they occasionally had visitors. No manual monitoring of data was performed during the experiments.

We evaluated the accuracy of the algorithms independently. Despite the low resolution and noise, the footstep segmentation accuracy was about 73 percent. Subjective evaluation of clips created using video and audio handover revealed natural camera changes with adequate sound levels. According to the results of a user study,[2] key frame extraction retrieved 80 percent of the most representative key frames. When evaluated with a data set of audio clips with a length of 200 minutes from each of the 25 microphones × 25 minutes, the sound-source localization algorithm had an average precision of 92 percent and a recall of 88 percent. The number of lighting change events detected per hour ranged from zero to 10. The system successfully detected all instances where lights were turned on or off, creating a brief summary of the events from the beginning until the end of the day. About 10 percent of the detected changes corresponded to insignificant events such as curtains blown by the wind.

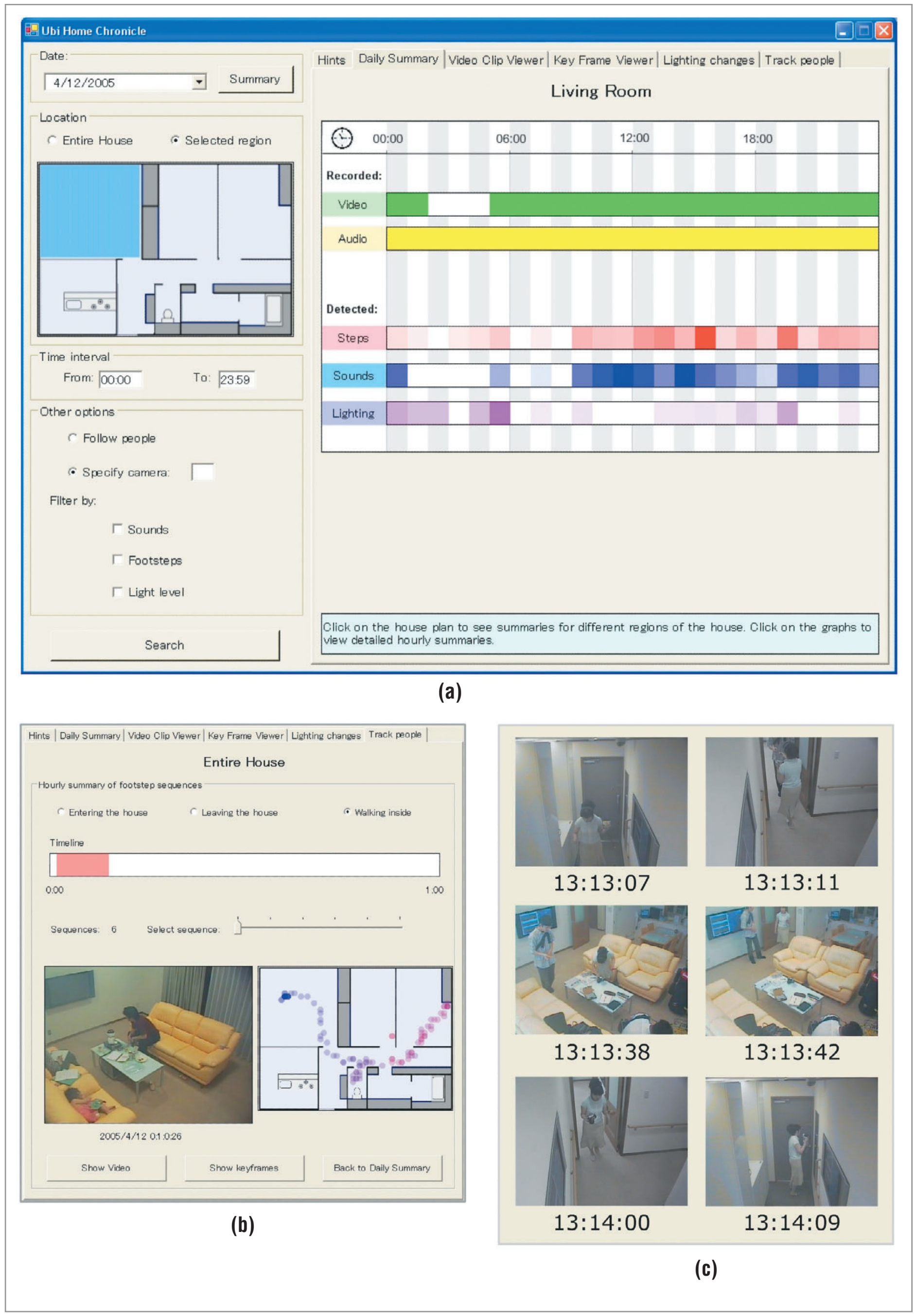

Figure 5. Event retrieval. (a) Activity pattern shows that the user went ot sleep around 1:00 a.m. (b) Retrieval of footstep sequences from midnight to 1:00 a.m. (c) Key frames can serve both as a summary of and an index for a video clip.

While the results suggest that the system has acceptable accuracy at the functional level, an overall evaluation in terms of usability and effectiveness is essential for a human-centered system. For this purpose, it was necessary to capture data from a real-life situation and use the system to retrieve user experience and satisfaction.

We thus designed and conducted a user study with the objectives of evaluating the system's performance and usability, identifying further requirements, and making improvements to the existing algorithms.

User requirements and usability study

During an early real-life experiment using a prototype system, a family of three—a married couple and their three-year-old daughter—stayed in the Ubiquitous Home for 12 days. Six months after their stay, we paid them to take part in a requirements and usability study/survey that took approximately three hours. Although this early prototype did not include sound-source localization and lighting-change detection, it was still satisfactory for our evaluation purposes as it contained the basic functionalities of a multimedia diary.

The study consisted of three parts. The first was a requirements study in which the subjects answered a questionnaire to specify what they would expect from a system for retrieving experiences at home. To ensure that the existing system's functionality did not influence user requirements, this section of the experiment was conducted before demonstrating the system.

The requirements study showed that the system already matched most of the specifications the subjects had in mind before using it.

In the second part, the subjects used the system for retrieval of experiences from a six-hour period during their stay in the Ubiquitous Home. This amounted to a total of 102 hours of video data and 150 hours of audio data.

The subjects were given a brief demonstration on how to use the system and were allowed to use it themselves. The authors were available in case the subjects needed advice, but they were not involved in using the system. The subjects were asked to select video clips they would like to keep, both as a motivational factor and to determine what kind of experiences generate interest in keeping a permanent record.

After using the system, the subjects rated its usability by answering a brief questionnaire based on the guidelines established by John Chin and colleagues.[4]

In the third part of the experiment, the subjects provided descriptive feedback about the system and were given the opportunity to suggest what additional requirements they would propose.

Results

The information these particular residents wished to retrieve from the media captured at home included the following:

- Things I did.
- How my child was playing when she was alone.
- Things I have forgotten.
- Things we did together.
- A summary of what I did each day.

Asked about the places and times media should be recorded, participants stated that they preferred continuous recording of data in common areas of the house and in the child's room.

For the usability assessment, the two parents provided the following responses using 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

Providing descriptive feedback about the system, the subjects stated that they had nearly forgotten what happened during that six-hour period, but they could recall almost everything that happened after using the system for about 40 minutes. They also discovered several things they didn't know, such as that the child consumed more food on her own sometime after breakfast. The subjects believed that the system would be useful to them as a family diary, recording events in their child's life, and helping to care for elderly family members.

The subjects 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 a video clip created when the child woke up in the morning, found that she was alone when she went into the living room, and ran to find her mother. The other type corresponded to activities they did together, such as eating meals and playing with the child. The family requested copies of both of these types of video clips. The subjects stated that they enjoyed using the system, and it was somewhat difficult to get them to stop using it to answer the feedback questionnaire.

An important observation is that the subjects did not feel uncomfortable about their life at home being recorded. The cameras and microphones were not hidden, and the family was aware of where the sensors were located. They stated that they got used to having cameras and sensors around them within the first couple of days.

The responses to the requirements study show that the system already matched most of the specifications the subjects had in 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 discover a few things they were not aware of. For example, the father stated, "I did not know that I spend so much time with my child. I always thought I had little time with her." He also observed, "I found that I was doing some office work during the day. I thought I did not work from home during that stay."

SYSTEM CHARACTERISTICS

The main difficulty in continuous capture is the large amount of disk space it consumes. For faster access, the video data is stored as frames and the audio as 1-minute clips, resulting in low compression of this data. Improved compression and storage techniques will be necessary to facilitate continuous data acquisition over a long period of time.

The floor sensors facilitate tracking people with less computational effort compared to using image analysis, which requires calibration and occlusion handling to achieve similar accuracy. However, movement of furniture causes superfluous data, making tracking difficult.

We conducted a pilot study of audio classification after sound-source localization. It was possible to classify audio into eight categories including voices, footsteps, and noise with 83 percent accuracy. We are working on integrating audio classification into the system.

Although the number of the subjects who took part in the user study is small, their keen interest in using the system and positive feedback justify continuing this work. The responses also provide valuable insights to identify further requirements and possible improvements. Further study with other families during ongoing development will help this evolve into a useful system.

The data was recorded only in the public areas of the house. Although this reduces the system's ability to function as a memory assistant, it is necessary as individual privacy is important even among members of the same family. Further, the system was helpful to the residents even with restrictions in locations. One reason for the success of the real-life experiment is that the residents were not confined to the house, which contributed to their enjoyment of their stay and the retrieval of their experiences.

We have implemented video retrieval and summarization for a home by analyzing signals from pressure-based sensors mounted on the floor. Hierarchical clustering followed by audio and video handover enabled the creation of personalized video clips using a large number of cameras.

An adaptive algorithm enabled retrieval of more than 80 percent of the key frames required for a complete and compact summary of the video. Basic audio analysis enabled accurate audio segmentation and source localization, thereby enhancing retrieval. The interface allowed users to incorporate their knowledge into the search process and obtain more accurate results for their queries. The residents who evaluated the system found it useful, enjoyed using it, and provided valuable feedback for improving the system.

Future work will focus on fusing the data from multiple modalities to enable detection and recognition of higher-level actions and events such as conversations, thereby enhancing the retrieval system's functionality. Face detection in retrieved images and video would provide additional information for searching within the data. Comments and feedback from the residents provide important information, as usability is an essential criterion for an effective retrieval system.

References

1. T. Yamazaki, "Ubiquitous Home: Real-Life Testbed for Home Context-Aware Service," *Proc. Tridentcom2005*, IEEE Press, 2005, pp. 54-59.
2. G.C. de Silva, T. Yamasaki, and K. Aizawa, "Evaluation of Video Summarization for a Large Number of Cameras in Ubiquitous Home," *Proc. ACM Multimedia 2005*, ACM Press, 2005, pp. 820-828.
3. G.C. de Silva, T. Yamasaki, and K. Aizawa, "Interactive Experience Retrieval for Ubiquitous Home," to appear in *Proc. ACM CARPE 2006*, ACM Press, 2006, pp. 45-48.
4. J.P. Chin, V.A. Diehl, and K.L. Norman, "Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface," *Proc. ACM Conf. Human Factors in Computing Systems* (CHI 88), ACM Press, 1988, pp. 213-218.

Gamhewage C. de Silva *is a research fellow in the Intelligent Media Laboratory at the University of Tokyo. His research interests include multimedia analysis and retrieval, vision-based human-computer interaction, and new interfaces for musical expression. He received a PhD in frontier informatics from the University of Tokyo. He is a member of the IEEE Computer Society. Contact him at chamds@hal.t.u-tokyo.ac.jp.*

Toshihiko Yamasaki *is an assistant professor in the Department of Information and Communication Engineering at the University of Tokyo. His research interests include 3D video processing, 3D computer graphics, and analog VLSI design. He received a PhD in electronic engineering from the University of Tokyo. He is a member of the IEEE and the ACM. Contact him at yamasaki@hal.t.u-tokyo.ac.jp.*

Kiyoharu Aizawa *is a professor in the Department of Information and Communication Engineering at the University of Tokyo. His research interests are multimedia applications for wearable and ubiquitous environments such as life log and processing of 3D video. He received a Dr. Eng. in electrical engineering from the University of Tokyo. He is a member of the IEEE, the ACM, the IEICE, and the ITE. Contact him at aizawa@hal.t.u-tokyo.ac.jp.*