

Capture and Efficient Retrieval of Life Log

Kiyoharu Aizawa, Tetsuro Hori, Shinya Kawasaki, Takayuki Ishikawa

Department of Frontier Informatics,

The University of Tokyo

+81-3-5841-6651

{aizawa,t_hori, kawasaki,ishikawa}@hal.t.u-tokyo.ac.jp

ABSTRACT

In "Wearable computing" environments, digitization of personal experiences will be made possible by continuous recording using a wearable video camera. This could lead to "automatic life-log application". It is evident that the resulting amount of video content will be enormous. Accordingly, to retrieve and browse desired scenes, a vast quantity of video data must be organized using structural information. In this paper, we are developing a "context-based video retrieval system for life-log applications". This system can capture not only video and audio but also various sensor data and provides functions that make efficient video browsing and retrieval possible by using data from these sensors, some databases and various document data.

Keywords

life log, retrieval, context, wearable

INTRODUCTION

The custom of writing a diary is common all over the world. This fact shows that many people like to log their everyday lives. However, to write a complete diary, a person must recollect and note what was experienced without missing anything. For an ordinary person, this is impossible. It would be nice to have a secretary who observed your everyday life and wrote your diary for you.

In the future, a wearable computer may become such a secretary-agent. In this paper, we aim at the development of a "life-log agent" (that operates on a wearable computer). The life-log agent logs our everyday life on storage devices instead of paper, using multimedia such as a small camera instead of a pencil.

There have been works to log a person's life in the area of mobile computing, wearable computing, video retrieval and database [1,2,3,8,9,10,11]. A person's experiences or activities have been captured from many different points of view. In one of the earliest works [7], various personal activities were recorded such as personal location and

encounters with others, file exchange, workstation activities, etc. Diary recording using additional sensors have been attempted in the wearable computing area. For example, in [2], a person's skin conductivities were captured for video retrieval keys. In [11], not only wearable sensors, but also RFIDs for object identification were utilized. Meetings were also recorded using sensors for speaker identification [9]. In database area, Mylifebits project attempts to exhaustively records a person's activities such as document processing, web browsing etc.

We focus on continuous capturing our experiences by wearable sensors including a camera. In our previous works [4,5], we used a person's brain waves and motion to retrieve videos. In this paper, we describe our latest work, which is able to retrieve using more contexts.

PROBLEMS IN BROWSING LIFE-LOG VIDEO

A life-log video can be captured using a small wearable camera with a field of view equivalent to the user's field of view. Videos are the most important contents of life-log data. By continuously capturing life-log videos, personal experiences of everyday life can be recorded by video, which is a most popular medium. Instead of writing a diary, a person can simply order the life-log agent to start capturing a life-log video at the beginning of every day. For a conventional written diary, a person can look back on a year at its end by reading the diary, and will soon finish reading the diary and will easily review events in the year.

However, watching life-log videos is a critical problem. It would take another year to watch the entire life-log video for one year. Then, although it is surely necessary to digest or edit life-log videos, editing takes even more time. It is the most important to be able to process a vast quantity of video data automatically.

Conventional Video Retrieval Systems

Recently, a variety of systems for video retrieval has been existing. Conventional systems take content-based approach. They digest or edit videos by processing the various features grasped from image or audio signals. For example, they may utilize color histograms extracted from image signals. However, even if they utilize such information, computers do not understand the contents of the videos, and they can seldom help their users to easily retrieve and browse the desired scenes in life-log videos. In

addition, such image signal processing requires very high computational costs.

Our Proposed Solution to this Problem

Life-log videos are captured by a user. Therefore, as the life-log video is captured, various data such as GPS, motion, etc. other than video and audio can be simultaneously recorded. By these information, computers may be able to use contexts as well as contents, thus, our approach is very different from conventional video retrieval technologies.

CAPTURING SYSTEM

The life-log agent is a system that can capture data from a wearable camera, a microphone and various sensors that show contexts. The sensors we used are a brain-wave analyzer, a GPS receiver, an acceleration sensor and a gyro sensor. All these sensors are attached to the notebook PC through, serial ports, USBs and PCMCIA slots. (figure 1 and figure 2)

Next, using a modem, the agent can connect into the Internet almost anywhere via the PHS (Personal Handy-phone System: Versatile cordless/mobile system developed in Japan.) network of NTT-DoCoMo. By referring to data on the Internet, the agent records "the present weather in the user's location", "various news on that day, which were offered by some news sites or some email magazines", "all web pages (*.html) that the user browses" and "all emails that the user transmitted and received".

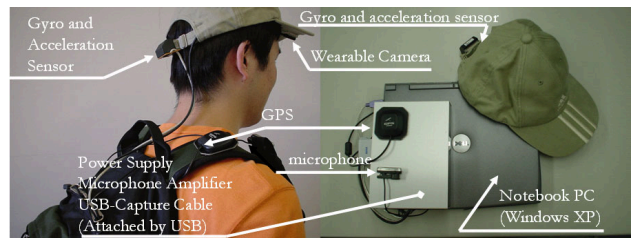


Figure 2. Capturing System

At last, the agent monitors and controls the following applications, "Microsoft Word", "Microsoft Excel", "Microsoft PowerPoint" and "Adobe Acrobat". In addition to web browsing and transmission and reception of emails, these applications are the main softwares used while people are using computer. Because of monitoring and controlling them, when the user opens document a file (*.doc; *.xls; *.ppt; *.pdf) of such applications, the agent can order each application to copy the file and save it as text data.

The user can use his cellular phone as a controller of operations "start/stop life-log". The agent recognizes the user's operations on his cellular phone via PHS.

RETRIEVAL OF LIFE-LOG VIDEO

We, human beings, save many experiences as a vast quantity of memories over many years of life while arranging and selecting them, and we can quickly retrieve and utilize necessary information from our memory. Some psychology researches say that we manage our memories based on contexts at the time. When we want to remember something, we can often use such contexts as keys, and recall the memories by associating them with these keys.

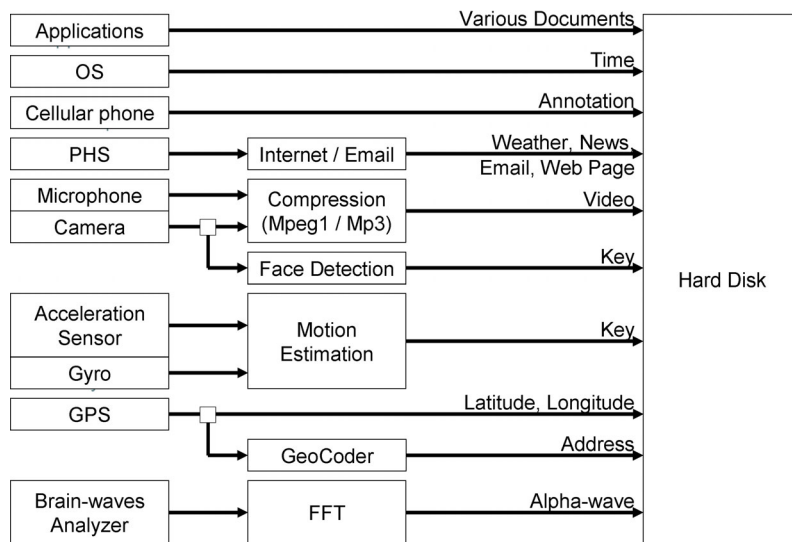


Figure 1. Diagram of Capturing System

For example, to recollect the scenes of a conversation, the typical keys used in the memory recollection process are such context information as "what, where, with whom, when, how"

A user may put the following query (Query A). "On a cloudy day in mid-May when the Lower House general election was held, after making my presentation about life-log, I was called to Shinjuku by the email from Kenji, and I talked with him while walking at a department store in Shinjuku. The conversation was very interesting! I want to see the scene to remember the contents of the conversation". In conventional video retrieval the low-level features of image and audio signals of the videos are used as keys for retrieval. Probably, they will not be suitable for queries compatible with the way we query to our memories as in Query A. However, data from the brain-wave analyzer, the GPS receiver, the acceleration sensor, and the gyro sensor correlate highly with the user's contexts. The life-log agent estimates its user's contexts from these sensor data and some database, and uses them as keys for video retrieval. Thus, the agent retrieves life-log videos by imitating the way a person recollects experiences from his memories. It is conceivable that by using such context information, the agent can produce more accurate retrieval results than by using only audiovisual data. Moreover, each input from these sensors is a one-dimensional signal, and the computational cost for processing them is low.

Keys Obtained from Brain-Wave Data

A sub-band [8-12 Hz] of brain waves is named α wave and it clearly shows the person's arousal status. When α wave is low [α -blocking], the person is in arousal, or in other words, is interested in or pays attention to something. We demonstrated that we can effectively retrieve a scene of interest to him using a person's brain waves in [4]. In Query A, the conversation was very interesting.

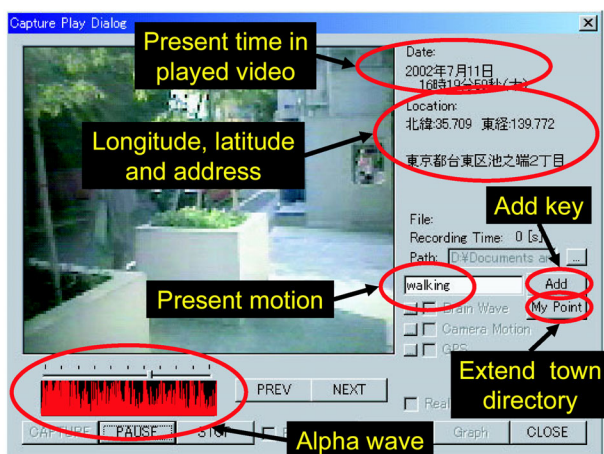


Figure 3. Interface for playing the video

Keys Obtained from Motion Data

The life-log agent inputs the data of the acceleration sensor and the gyro sensor to the K-Means method and HMM and estimates the user's motion state. The details are in our previous paper[5]. In Query A, the conversation was held while the user was walking.

Keys Obtained from Face Detection

The life-log agent detects a person's face in life-log videos by processing the color histogram of the video image. Our method only uses very easy processing of the color histogram. Accordingly, even if there is no person in the image, when skin color is predominantly included, the agent make a wrong detection. But, the agent shows its user the frame images and the time of the scene in which the face was detected. If it is a wrong detection, the user can ignore it and can also delete it. If the image is detected correctly, the user can look at it and judge who it is. Therefore, identification of a face is unnecessary and simple detection is enough here. In Query A, the conversation was held when the user was with Kenji.

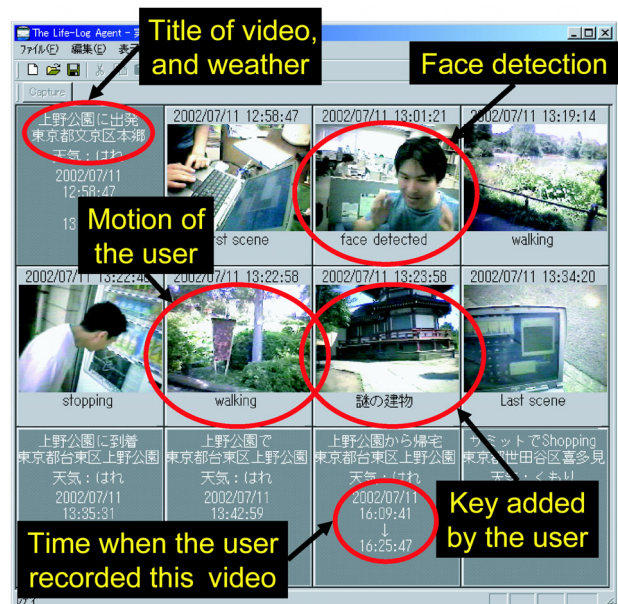


Figure 4. Interface for managing videos

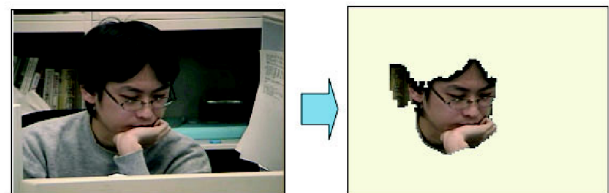


Figure 5. A result of face detection

Keys Obtained from GPS Data

From the GPS signal, the life-log agent acquires information about the position of its user as longitude and latitude when capturing a life-log video. The contents of

videos and the location information are automatically associated. Longitude and latitude information are one-dimensional numerical data that identify positions on the Earth's surface relative to a datum position. Therefore, they are not intuitively readable for users. However, the agent can convert longitude and latitude into addresses with hierarchical structure using a special database, for example, "7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan". The results are information familiar to us, and we use them as keys for video retrieval.

Latitude and longitude information also become information that we can intuitively understand by being plotted on a map as the footprints of the user, and thus become keys for video retrieval. "What did I do when capturing the life-log video?" A user may be able to recollect it by seeing his footprints. The agent draws the user's footprint in the video under playback using a thick light-blue line, and draws other footprints using thin blue lines on the map. By simply "dragging his mouse" on the map, the user can change the area displayed on the map. The user can also order the map to display the other area by clicking arbitrary addresses of all the places where footprints were recorded. The user can watch the desired scenes by choosing arbitrary points of footprints.

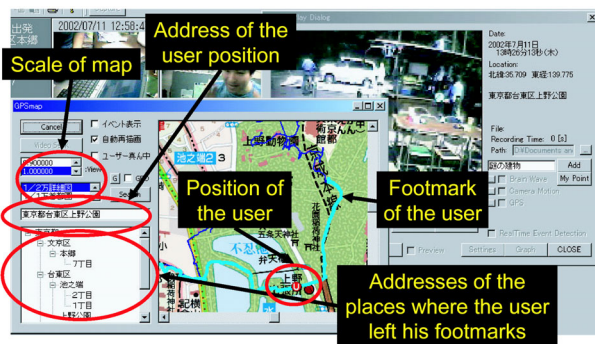


Figure 6. Interface for retrieval using a map

Moreover, the agent has a town directory database. The database has a vast amount of information about one million or more public institutions, stores, companies, restaurants, and so on, in Japan. Except for individual dwellings, the database covers almost all places in Japan including small shops or small companies that individuals manage. In the database, each site has information about its name, its address, its telephone number, and its category with layered structures.

Using this database, a user can retrieve his life-log videos as follows. He can enter the name of a store or an institution, or can input the category. He can also enter the both. For example, we assume that the user wants to review the scene in which he visited the supermarket called "Shop A", and enters the category-keyword "supermarket". To filter retrieval results, the user can also enter the rough location of Shop A, for example, "Shinjuku-ku, Tokyo".

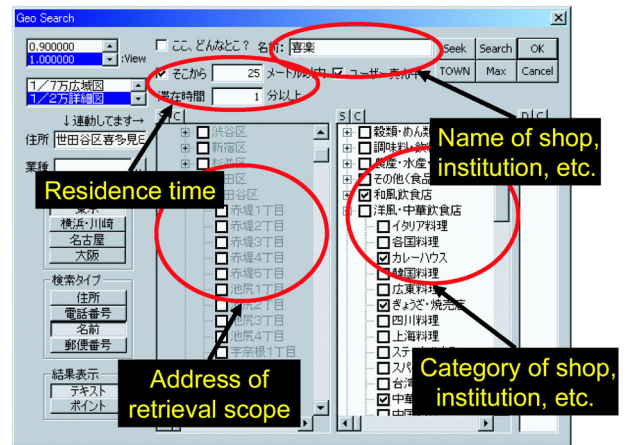


Figure 7. Retrieval using the town directory

Because the locations of all the supermarkets visited must be indicated in the town directory database, the agent accesses the town directory, and finds one or more supermarkets near his footprints including Shop A. The agent then shows the user the formal names of all the supermarkets which he visited and the time of visits as retrieval results. Probably he chooses Shop A from the results. Finally, the agent knows the time of the visit to Shop A, and displays the desired scene. In Query A, the conversation was held at a shopping center in Shinjuku.

The agent may make mistakes, for example, to the query shown above. Even if the user has not actually been into Shop A but has passed in front of it, the agent will enumerate that event as one of the retrieval results.

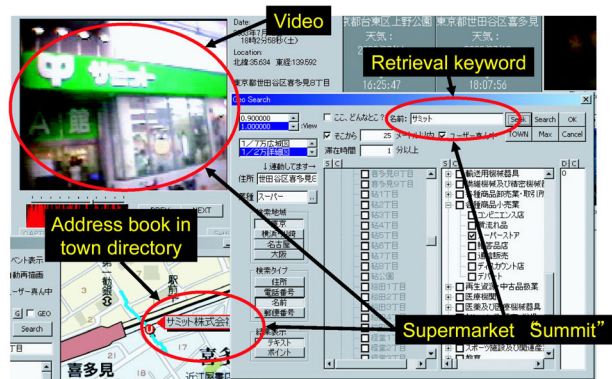


Figure 8 Retrieval experiments

To cope with this problem, the agent investigates whether the GPS signal was received during the event. If the GPS became unreceivable, it is likely that the user went into Shop A. The agent investigates the length of the period when the GPS was unreceivable, and equates that to the

time spent in Shop A. If the GPS did not become unreceivable, the user most likely did not go into Shop A.

We examined the validity of this retrieval technique. First, we went to Ueno Zoological Gardens, the supermarket "Summit", and the drug store "Matsumoto-Kiyoshi". We found that this technique was very effective! For example, when we referred to a name-keyword "Summit", we found the scene that was captured when the user was just about to enter "Summit" as the result. When we referred to the category-keyword "drug store", we found the scene that was captured when the user was just about to enter "Matsumoto-Kiyoshi", and similarly for Ueno Zoological Gardens. These retrievals were quickly completed; retrieval from videos for three-hours took less than one second.

Keys Obtained from Time Data

The agent records the time by asking the operating system for the present time, and associates contents of life-log videos with the time when they were captured. In Query A, the conversation was held in mid-May.

Keys Obtained from the Internet

The life-log agent records the weather and news on that day, web pages that the user browses and emails that the user transmitted and received. These data are automatically associated with time data. Afterwards, these data can be used as keys for life-log videos retrieval. In Query A, the conversation was held after the user received the email from Kenji on a cloudy day when the Lower House general election was held.

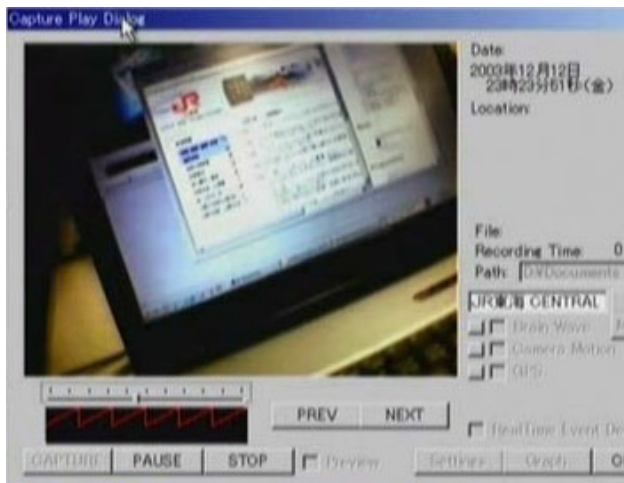


Figure 9. A result of retrieval from Web-document

Keys Obtained from Various Applications

All the document files (*.doc; *.xls; *.ppt; *.pdf) that user opens are copied and saved as text. These copied document files and text data are automatically associated with time data. Afterwards, these text data can be used as keys for life-log videos retrieval. In Query A, the conversation was

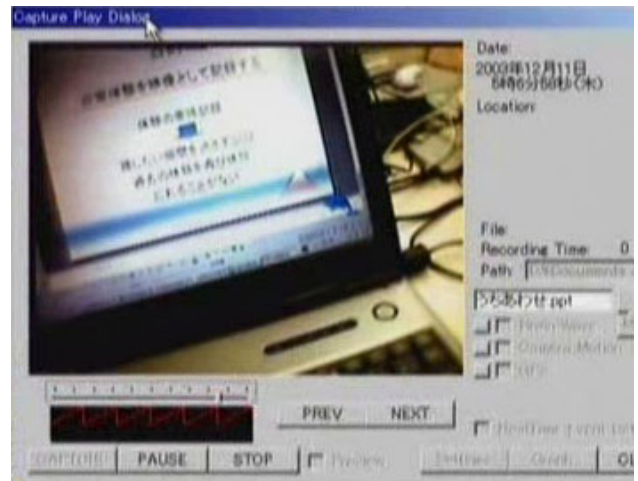


Figure10. A result of retrieval from PowerPoint-document

held after the user made presentation about life-log. (assume that PowerPoint was used at his presentation.)

Reversely, the agent can also perform video-based retrieval for such documents including web pages and emails.

Keys Added by the User

The user can order the life-log agent to add retrieval keys (annotation) with an arbitrary name by simple operations on his cellular phone while the agent is capturing a life-log video. This enables the agent to identify a scene that the user wants to remember throughout his life, and thus the user can access easily to the videos that were captured during precious experiences.

Retrieval with a Combination of Keys

Consider Query A again. The user may have met Kenji many times during some period of time. The user may have gone to a shopping center many times during the period. The user may have made presentation about life-log many times during the period...etc.

Accordingly, if a user uses only one kind of key among the various kinds of keys when retrieving life-log videos, too many results which he does not desire will appear. By using as many different keys as possible, only the desired result may be obtained, or at least most of the undesired results can be eliminated.

CONCLUSION

By using the data acquired from various sources while capturing videos and combining these data with data from some databases, the agent can estimate its user's various contexts with high accuracy and high speed that do not seem achievable with conventional methods. These are the reasons the agent can respond to video retrieval queries of various forms correctly and flexibly.

REFERENCES

1. S.Mann, 'WearCam' (The Wearable Camera), In *Proc. of ISWC 1998*, 124-131.
2. J.Healey, R.W.Picard, A Cybernetic Wearable Camera, In *Proc. of ISWC 1998*, 42-49.
3. J. Gemmell, G. Bell, R. Lueder, S. Drucker, C. Wong, MyLifeBits: fulfilling the Memex vision, In *Proc. of ACM Multimedia 2002*, 235-238.
4. K.Aizawa, K.Ishijima, M.Shiina, Summarizing Wearable Video, In *Proc. of IEEE ICIP 2001*, 398-401.
5. Y.Sawahata, K.Aizawa, Wearable Imaging System for Summarizing Personal Experiences, In *Proc. of IEEE ICME 2003*.
6. T.Hori, K.Aizawa, Context-based Video Retrieval System for the Life-log Applications, In *Proc. of MIR 2003*, ACM, 31-38.
7. M.Lamming and M.Flynn, Forget-me-not: intimate computing in human memory, In *Proc. FRIEND'21*, Int. Symp. Next Generation Human Interface, Feb.1994
8. B.J.Rhodes, The wearable remembrance agent: a system for augmented memory, In *Proc. of ISWC 1997*
9. N.Kern et al., Wearable sensing to annotate meeting recordings, In *Proc. of ISWC 2002*
10. A.Dey et al., The conference assistant : combining context-awareness with wearable computing, In *Proc. of ISWC 1999*
11. T.Kawamura, Y.Kono, M.Kidode, Wearable interface for a video diary: towards memory retrieval, exchange and transportation, In *Proc. of ISWC 2002*

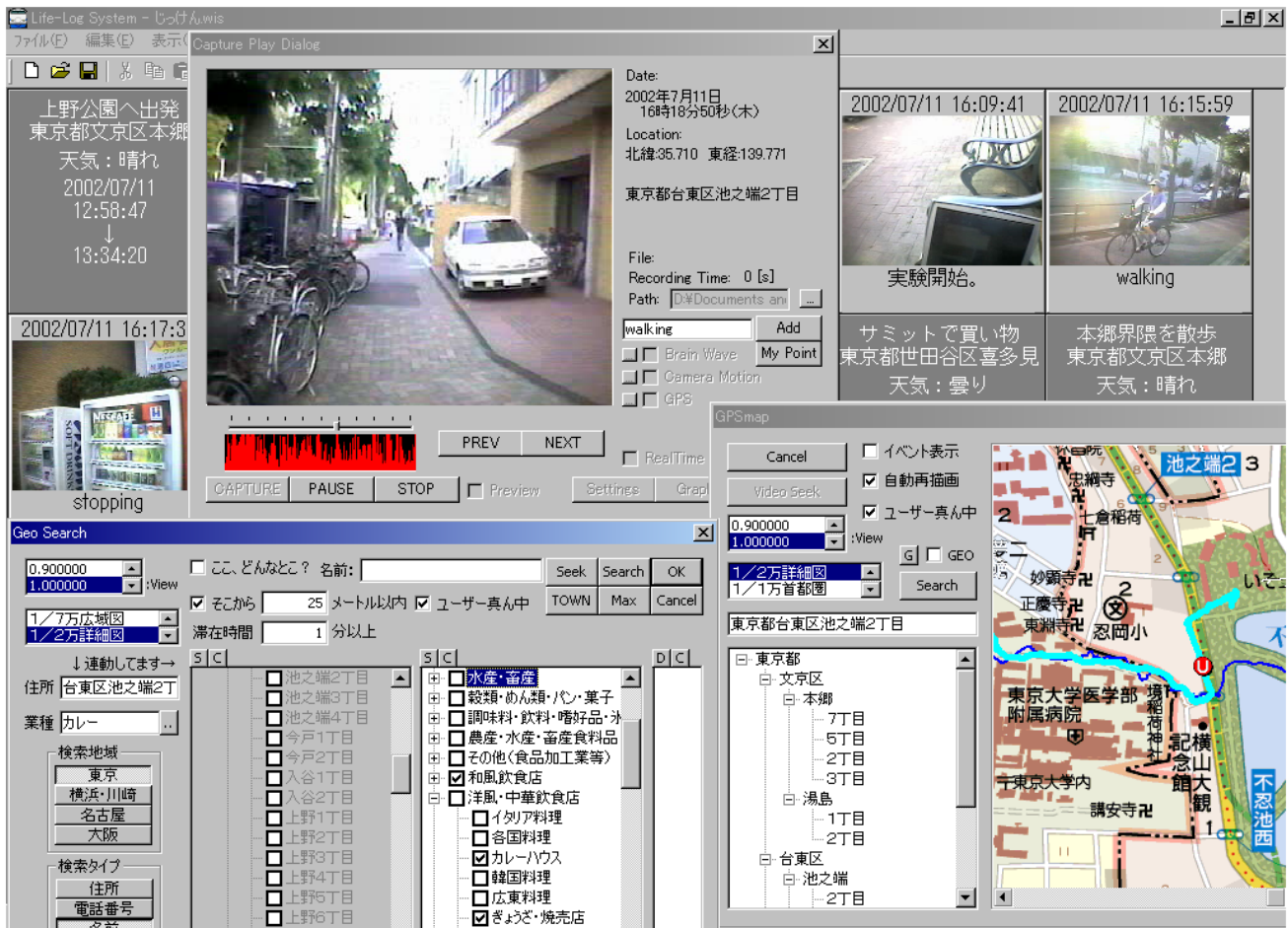


Figure 11. Interface of the life-log agent for browsing and retrieving life-log videos