SenseWeb : Browsing Large Multimedia Collections in a Multi-user Interaction Environment

Roberto Lopez-Gulliver, Tomohiro Sato, Hiroko Tochigi, Masami Suzuki
ATR Media Information Science Research Laboratories
2-2-2 Hikari-dai Keihanna Science City Kyoto 619-0288, JAPAN
{gulliver,tomohiro,tochigi,msuzuki}@atr.jp

Abstract

This paper introduces a hands-free multi-user interaction environment for browsing large multimedia collections. The system allows multiple users to simultaneously browse, select and refine the search of images in large collections, in a collaborative way. Its touch-based interface provides a natural interface to interact with multimedia contents by using simple hand gestures. Experiment results are discussed to confirm the suitability of the system's visualization technique as well as effectiveness of simultaneous user interaction for image classification.

1. Introduction

We are now experimenting probably the most data rich period since the advent of computers. Advances in data collection and information retrieval techniques have allowed us to digitize, organize and classify large amounts of multimedia information [8,9].

A plethora of novel visualization and user interaction techniques have been proposed to make an effective browsing of such multimedia information [1,2]. Most of these techniques are designed for single-users in a desktop interface with mouse and keyboard as input devices.

However, in order to provide more intuitive browsing experiences and support collaboration between users simple but effective interaction techniques as well as a multi-user interaction environment are required, allowing users to simultaneously interact with the data at hand [3,4,5,6]

This paper introduces a touch-based multi-user interaction environment, shown in Fig. 1, that allows users to collaboratively browse, interact and visualize large amount of images and sounds from large multimedia collections. Thus providing users with intuitive collaboration capabilities not found in traditional single-user, mouse-keyboard based systems.

The following sections describe the proposed multi-user interaction environment in terms of visualization and multi-user capabilities, with discussions of results from early experiments.

Figure 1. **SenseWeb: A touch-based multi-user interaction environment.**

2. The SenseWeb System

2. 1 Setup and Modules Overview

The SenseWeb system aims at providing multi-user capabilities and intuitive interaction paradigms using multimedia data, to support users in discussions, brain-storming or collaboration situations. To this end, a touch-based multi-user environment was designed and prototyped.

Figure 2 shows the hardware setup of the system. It consists of a 150-inch rear-projection screen, a video camera with infrared(IR) paper filter and halogen lamps as IR light sources illuminating the screen. An optional set of speakers for sound browsing as well as a microphone for commands and keyword input is also provided.

By analyzing the IR shadows of the users' hands seen from the video camera, the system is able to track the users' hands positions thus turning the large screen into an interactive touch-based multi-user screen.

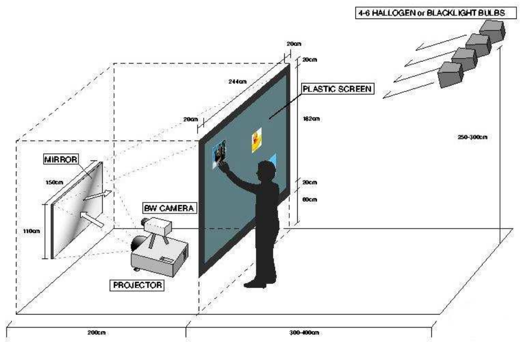

Figure 2. SenseWeb hardware setup: 150-inch screen, rear projection, camera with infrared (IR) filter and halogen lamps (as IR light source)

The system consists of four modules: speech input, hand tracking, image/sound search servers and graphics visualization module. They all can run as standalone modules on remote or local machines, communicating via sockets or shared memory respectively.

2.2. Image Browsing Application

The very first application on top of the SenseWeb multi-user interaction environment was a collaborative image/sound browsing application. We decided to use the Internet as large a database of images and sounds although the same algorithms also work with local databases.

In this application, mouse and keyboard are replaced by hand-touching and speech. Multiple users can browse and interact with images and sounds from the Internet, or the local database if needed, by simply speaking keywords to the microphone. These keywords were recognized and sent as requests for images/sound related to them from the Internet. Users can then touch the keyword icons (mother icons) so generated to trigger the downloading of images/sound related to them. These newly downloaded image icons (child icons) would start to fly outwards with constant velocity in random radial directions with its center in the originating mother icon.

By using their bare hands multiple users can interact with the images' icons by using one or two hands. Figure 3 schematically shows the browsing interaction primitives available. For example, by using two hands or in collaboration with other user, dragging together to mother icons users can refine the search to the logical AND of the two keywords attached to them. By touching the icons with both hands users can zoom up the icons for a closer look and by holding them this way for a certain time, 3-5 seconds, they can send them

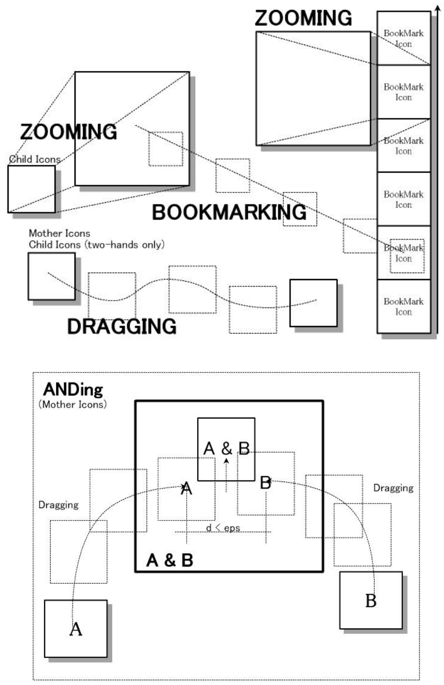

***Figure 3*. Browsing Interaction Primitives: touch, zoom in, bookmark, drag, and |"AND" with both hands (bottom)**

to the bookmark area in the right part of the screen, as shown in Fig 1 and 3.

3. Evaluation Experiments

In order to assess the suitability of our visualization method and the effectiveness of the multi-user capabilities of the SenseWeb system we carried out two experiments. We briefly describe the experiments and discuss the results here. More details can be found in our papers [6,7] respectively.

3.1. Visualization

The goal of the experiment is to compare the effectiveness, by measuring response time, of the SenseWeb visualization method versus the generic Internet image browser visualizations.

The experiment task consists on searching for a target image within a 100-image data set. Each user is asked to perform this task with 10 different data sets for each of the two visualization methods. Response times were recorded for later analysis. The number of subjects was 30. Under the experiment setup, in the generic Internet image browser visualization method users are presented with rows-columns display. Users had to scroll down and turn pages to search for the target image. In the SenseWeb visualization method users are

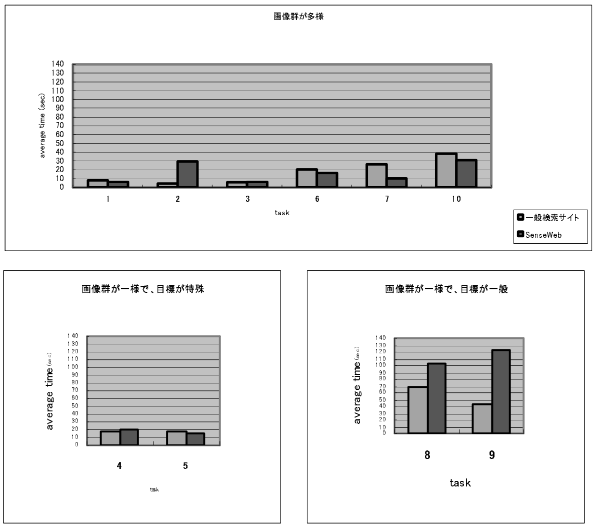

Figure 4. Average response times (in secs) for image search using the SenseWeb system (red) and a generic Internet browser layout (blue).

presented with the fireworks-like fashion visualization method, mother and child icons, described in the previous section.

Results of average response time, in secs, for both visualization methods are shown in Fig. 5. The SenseWeb visualization method performed extremely well in 8 of the 10 data sets, specially on the ones where the target images had very distinctive features, like color, texture, motif, etc with respect to the whole image set. This confirmed our expectation that having the SenseWeb system bring the data and animate the images by moving them is more effective than having the user to scroll and turn pages.

3.2. Multi-user Collaboration

The goal of the experiment was to assess the effectiveness, by measuring response time, of the multi-user (multi-point) capabilities of the SenseWeb system versus its single-user(single-point) version. The task consisted in two users collaboratively browsing and selecting 8 images from a 100-image data set. Each pair of users is asked to perform this task with 3 different data sets for each of the two interaction modes, that is single-point versus multi-point modes.

The number of subjects was 20, 10 pairs, male or female, university students. In order to force users to touch the screen as much as possible, we prepared corresponding monochrome versions of the color images. Color images would be active and remain so after being touched once. Users could drag and/or throw the image icons for discussion or put them in the so called "selection row". Icons already in the selection row can be dragged out or interchanged with other if necessary.

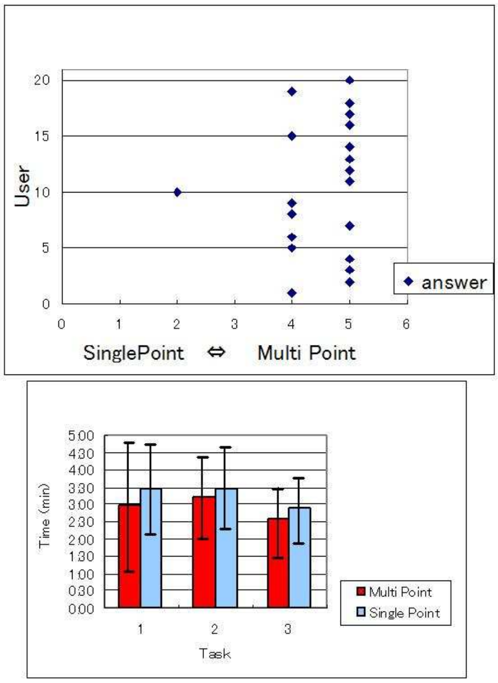

Figure 5. Single and multi-point results: Ease of use (top) and average task completion times (bottom).

The task is considered finish once the two users agree on the 8 selected images. After finishing with the 6 tasks both users are asked to answer a questionnaire comparing the modes, single versus multi-point, in terms of usability and user satisfaction.

In the multi-point mode users can simultaneously manipulate the images regardless of the other user's current interaction. In the single-point mode, only one user at the time can manipulate the images, taking turns if needed. The one who touches the screen first takes precedence over any other touches.

Figure 5 (top) shows the subjects' answer to the question "Which interaction mode you fell easier to use" within a range of 1 (single-point) and 5 (multi-point). All the subjects, except for one, regarded the multi-point mode easier to use. Most users pointed out the advantages of the multi-point mode by allowing them to switch working in a parallel(asynchronous) or in a serial(synchronous) way.

Figure 5 (bottom) shows the subjects' task completion times for both interaction modes, averaged over all the users. Also here the multi-point interaction mode proved to be slightly faster in task completion times that its single-point counterpart.

4. Discussion

Even though the two experiments, evaluating the visualization method and the multi-point interaction mode, confirmed the effectiveness of them, more detailed experiments setup and different tasks with a different levels of granularity should be done. By simple measuring task completion time in the multi-point versus single-point interaction points tells us nothing about what the actual communication and collaboration between the subjects. A more rigorous analysis of the videotaped sessions could tell us more about how and when switched between parallel and serial modes of collaboration.

In the visualization evaluation, we should be also doing experiments by varying not only the visualization method, but also by varying the different parameters of the SenseWeb visualization method, that is, velocity of the child icons, “birth-rate”, size of the icons, the total number of icons displayed on the screen at any one time, and many others.

At least we are now confident that our visualization method as well as our multi-point interaction mode are actually effective when browsing large multimedia collections.

5. Conclusions and future directions

We have presented a multi-user interaction environment for browsing large multimedia collections that effectively allows users to collaborate. The main features of the systems are its easy-to-use touch-based interface and its multi-user capabilities. Preliminary experimental results show the effectiveness of its multi-user capability as well as its visualization method in supporting collaborative tasks.

In the near future we plan to have the SenseWeb system serve as an environment to share experiences among users from multimedia data collected from existing databases or from ubiquitous sensors. The large amount of data available in this scenarios could serve as a good test bed for the SenseWeb system.

We are also planning to design and prototype a whole array of applications, ranging from games, to edutainment to discussion support, making use of the multi-user capabilities of the system.

There are also plans to have several SenseWeb system networked to support collaboration between groups.

Acknowledgments

This research was supported by the National Institute of Information and Communications Technology of Japan.

References

[1] Y. Abe, A. Sakamoto, T. Kawamura, Y. Kihara, "A Study of Assisting method for Web space Cruising in InfoLead", IPSJ SIG Technical Report, No.48-007, Japan, May 2003, pp. 37-42.

[2] H. Tobita, D. Iso, J. Rekimoto, "VelvetPath: Layout Design System with Sketch and Paint Manipulations", IPSJ Journal, Vol.44 No.11, Japan, Nov 2003, pp. 2528-2537.

[3] J. Rekimoto, "SmartSkin: An infrastructure for free-hand manipulation on interactive surfaces", In Proc of CHI 2002, 2002, pp. 113-120.

[4] Shen, C., Everitt, K.M., Ryall, K., "UbiTable: Impromptu Face-to-Face Collaboration on Horizontal Interactive Surfaces", ACM Intl. Conf. on Ubiquitous Computing, Oct. 2003.

[5] R. Lopez-Gulliver, T.Satoh, H. Tochigi, M. Suzuki, N.Hagita, ”SenseWeb: Collaborative Image Classification in a Multi-User Interaction Environment”, *ACM Multimedia*, New York, 2004.10 (to appear)

[6] H.Tochigi, R. Lopez-Gulliver, T. Sato, M.Suzuki, "Evaluation of a Cooperative Task for Image Classification using the SenseWeb Multi-User Environment", *IPSJ Technical Report*, 2004-GN-51, pp.115-120, 2004.03 (in Japanese)

[7] T. Satoh, R. Lopez-Gulliver, H. Tochigi, M. Suzuki, "Visualizing Large Amount of Images: Evaluation of the Visualization Method used in the SenseWeb Interaction Environment", In Proc. of Visual Information Research Assoc. of IPSJ, Kyoto Japan, Sept 2003, pp. 24-27

[8] Han, J and Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2000.

[9] G. Salton, Introduction to Modern Information Retrieval, New York, McGraw Hill, 1983.