SenseWeb: A Multi-user Environment for Browsing Images from the Internet

Roberto Lopez-Gulliver, Norihiro Hagita, Masami Suzuki, Tomohiro Satoh, Hiroko Tochigi

Abstract
This paper proposes a novel touch-based multi-user interaction environment for browsing images from the Internet. A large touchable screen serves as a shared data space where multiple users simultaneously interact with Internet images by using simple hand movements and gestures. Users can select, bookmark, clean up and refine the search of the incoming images. The system provides a natural and easy-to-use interface to browse and select items from large databases, with uses varying from discussion and brain storming support to edutainment and fun. Early preliminary results show the suitability of our image interactive visualization method for browsing large image collections.

1. Introduction
Advances in data collection have flooded us with data, creating an urgent need for new techniques and tools that can intelligently assist in transforming this data into useful knowledge [7].

On the other hand, interaction with Internet multimedia data, images and sound, is still mostly confined to the standard desktop metaphor of keyboard and mouse through the current standard representations found in generic Internet image browsers. While searches employing these browsers are certainly very effective, they are however not optimal for intuitive and collaborative browsing experiences. Especially when the user does not yet know what he is looking for, or he just wants to immerse himself playfully in the available data on the Internet.

In order to provide intuitive and collaborative browsing experiences, a hands-free and multi-user enabled interaction environments are necessary [2, 3, 8]. User interaction and visualization techniques work together to provide better ways of browsing and selecting items from large multimedia collections [1, 5, 6] with their usage varying depending on the task at hand.

This paper introduces a large hands-free and multi-

Figure 1. SenseWeb: A large hands-free multi-user interaction environment.

user enabled interaction environment, called SenseWeb, refer to Fig. 1, to browse and interactively visualize large image collections directly found in the Internet. These features enable intuitive, ease-to-use interaction and collaboration capabilities that could not be achieved with the traditional mouse-keyboard based interactions. Preliminary results, about the proposed visualization method compared with generic Internet image browsers, are also discussed.

2. System Architecture
2.1. Hardware Setup
The system consists of a 150-inch rear-projection acrylic screen. A black and white video camera, with a paper infrared (IR) filter, behind the screen and positioned so its field of view covers the screen completely. An array of black light bulbs are placed away in front and above illuminating the screen with very low visible light, but with high levels of infrared (IR) light. See Figure 2.

2.2. System Overview
The system is composed of four modules: speech input, hand tracking, image search server and graphics visualization module as shown in Figure 3. All the modules can run independently of each other and even on different machines. They communicate via sockets (remote) or shared memory (local) with proper read-write lock synchronization.

The speech input module captures and recognize the
user utterances and pass them as keywords, via shared memory, to the image search engine module to start the downloading of related images from the Internet. Up to twenty different Internet images search engines are used. It can also work with a local database if necessary. The images downloading for that keyword will continue forever unless there is a request to stop. Multiple requests are deal in a transparent way by tagging the keywords with ids and initiating a new thread process for the downloading thus keeping track of which image correspond to which keyword. A remarkable feature of the image search module is that it can spam a large number of threads for downloading without any performance and synchronization delays between them. The hands tracking module analyze the user’s infrared shadows from the video camera, threshold these images, finds the hands blob using an optimized seed-fill algorithm and outputs a list of normalized x, y coordinates and area size for each of the hand blobs detected. A notable feature of the hand-tracking module over traditional mouse-based methods is that it keeps track of the hands by labeling them with ids thus providing multi-user hand gesture recognition.

The graphics rendering module is the so-called application level module. It gets the downloaded images from shared memory for display. It also reads the user’s hands coordinates and uses them to simulate multiple mouse inputs mapping them into events appropriate for the application at hand. In this case, to start and stop image downloading as well as an array of interaction techniques for browsing the images, described in the following sections.

3. Visualization

3.1. Screen Layout (GUI)

After experiencing with many different screen layouts we choose the screen layout shown in Figure 4. The decision was based on a number of factors related to the problem at hand, that is, browsing and selecting from a large amount of images from the Internet collaboratively. We aimed at keeping it simple from the user point of view, deliberately avoiding complex menus thus complicating the user interaction.

One important factor on the simplicity of the design was that our main goal is to provide a wide range of users, with or without computer interaction experience, with an easy to use and intuitive environment to browse and interact with Internet images. By keeping the screen layout simple, users can concentrate on the interactive browsing and selection of the images rather than worrying about how to make the best use of graphical interface.

The screen layout, as shown in Figure 4, consists of a status bar at the bottom used to provide feedback on the keywords recognized by the speech input module. A bookmark column on the right side, to store selected items for further processing. And the main interaction window area at the center, where users interact with the images using their bare hands as the only input.

3.2 Visualization method

As part of the screen layout and interaction design, we choose to have only two different types of image icon representations. So called “mother” icons are generated for each recognized keyword from the speech input module, they then appears from the bottom of the screen and float upwards waiting to be touched by the users. Once touched they stop moving and trigger, a request to the image search module to search and download images from the Internet corresponding to that specific keyword. As soon as related images are downloaded they are displayed in a fireworks-like fashion by flying outwards with constant velocity in random radial directions with its center in
the originating mother icon. These icons are referred as “child” icons and when touched behave in different ways than its “mother” counterparts, refer to section 4.

In order to assess the suitability of our proposed visualization method we performed some preliminary experiments. The experiment was intended to compare the effectiveness, by measuring response time, of the visualization method compared with generic Internet image browser visualizations. Here, we only describe the experiment setup and the results, for a more detailed explanation refer to our paper [6].

The experiment task consisted in searching for a target image within a 100-image set. This task to be repeated for 10 different target images for each user. Experiment user subjects were 30 in number. In the first scenario users were presented to the generic rows-and-columns Internet image browser visualization display. Users had to scroll down and turn pages in order to search for the target image. In the other task setup, users were presented with the same images flowing from the center of the screen in a fireworks-like fashion as presented in the SenseWeb system, described above.

Results of the average search times are shown in Figure 5. The SenseWeb visualization method performed extremely well in most cases, especially on the ones where the target image had distinct features, like color, texture and appearance. We tried to have 2 different groups of image set, one with very similar features and other with distinctive features. The generic Internet image browser visualization method performed, in average, better than the SenseWeb one only in two of the ten image sets, concretely in the ones with images with very similar features. These results show that the SenseWeb visualization method is appropriate for browsing large image data sets interactively.

4. User Interaction

Our main goal was to ease the interactivity of the image browsing experience by providing hand-free, and intuitive interaction functionality. Ideally the user can simple speak a keyword describing the images he would like to browse and use his hand to touch and interact with them.

The users’ interaction flow allowing them to interact with the images is as follows. First, trigger the image downloading by saying a keyword, wait for the mother icons to appear and touch them to trigger the download of related images thus generating child icons from them. With very simple gestures, one hand versus two hands and the touch’s length of time, users can perform most of the image browsing tasks. Clean away the screen space from undesired images by simple touching them with one hand, this will increase the velocity of the child icon touched so it will go out of the screen faster. If the user is interested in any of the incoming images he can use his two hands to touch them, this will zoom up the image to view it in more detail. He can then drag it over to show it to the other users or to a position he desires. If he holds it long enough, a configurable parameter in our setup set to 3 secs, with his two hands the system guesses that he was interested in that particular image and send it to the bookmark column to the right where users can then zoom up the URL where that image was taken from.

One distinctive feature of the SenseWeb system is its multi-user capabilities. Users can use both hands at the same time or collaborate with other users to join two different mother icons, originated from two different user speech input keywords A and B, to further refine the search to their boolean AND, A&B, of the two keywords. All of the above mentioned user interaction functionalities are shown with actual user interaction sessions in Figure 6.
5. Discussion

Literally thousands of image search and visualization techniques have been researched and deployed in real world applications, varying in their user interaction schemes depending on the amount of images as well as the skills of the person to use them. Considering the main goal of the SenseWeb system to provide a user with an easy-to-use, intuitive and entertaining browsing experience, we had to make a compromise between the complexity of the interaction and visualization methods, and the user’s simple purpose of finding or merely discovering what is there in the Internet.

We believe that large image collections are better browsed by a group than by a single user. The SenseWeb multi-user capabilities allow for a collaborative browsing and interaction experience providing brainstorming and discussion sessions’ support. Even more, users can seamlessly switch between parallel (asynchronous) and collaborative (synchronous) interactive sessions. For example, we noticed that users start by simply bringing as many images as possible by their own but then switch to a group mode, sharing their findings with the other users with images of common interest. Figure 1 show tree users, two of them collaborating and the other interacting by his own. The two users on the left of the figure discuss one image of common interest by zooming it up, while the user on the right concentrates himself in a different image.

6. Conclusions and future directions

We have proposed an ease-to-use multi-user capable interaction environment for collaboratively browsing large image collections from the Internet, or local databases. The simplicity of the system’s user interaction modes and visualization techniques makes it appealing to a large audience of users with different computer skills.

We are now conducting user study experiments to assess the suitability of the system’s multi-user capabilities for image browsing.

Apart from investigating other types of user interaction we plan to use our system to visualize a large amount of user’s data from one of our ubiquitous computing projects. The amount of images, video and sensor data collected in this project make it a good challenge to test the usability of the system in real situations.

7. Acknowledgments

The research reported here was supported in part by a contract with the Telecommunications Advancement Organization of Japan entitled, “A Study of Innovational Interaction Media towards a Coming High Functioned Network Society”.

References