

A flexible image browsing interface tester: Design, implementation and preliminary results in a face recognition test.

Peter E. Jörgensen

School of Information Studies, Florida State University, Tallahassee, FL, 32306-2100,

Email: pjorgensen@lis.fsu.edu

This paper describes a software tool for exploring factors that affect image browsing. The tool, which can present images in storyboard or slideshow formats, can easily be configured to change the size of the images, the display rate, and other parameters of interest. Results of a preliminary experiment involving 26 subjects using the slideshow mode support earlier findings and are reported.

Introduction

Images have been fundamental for communication since before written language. The use of images for communication is experiencing a resurgence as a post-text period of video and translingual consumer products develops. A wide range of users now select images from digital collections, rather than creating them expressly for a use. The number of images from which one can choose is huge and growing rapidly. Collections on CD-ROM have grown to include thousands of images on each CD. The World Wide Web (WWW) has millions of images, many collected in huge catalogs by portals and search engines. Google, which claims to have the most comprehensive image search on the web, indexes over 425 million (Google, 2004). Improvements in image indexing, image query interfaces, presentation and the like, though significant during the past decade, have not been sufficient to eliminate the need to browse through a final collection of images to make a selection. The importance of the browsing interface makes a solid understanding of browsing capabilities fundamental to our design of good systems. Yet the literature yields little in the way of basic research regarding human image browsing performance. This paper attempts to begin to fill that need by presenting a flexible instrument which can be used to measure a variety of relationships between image size, bit depth, viewing duration and interface effectiveness. It also reports on the results of a user study investigating the effects of display duration on recognition accuracy.

Background

Two types of image retrieval systems are widely in use: text-based, which use keywords assigned in a variety of

ways to the images; and content-based (CBIR), which use algorithms to visually parse images. Neither, however, can currently provide reliable access to specific images based on object inclusion or semantic content (Berinstein, 1999). This is due to two very different reasons. Text-based indexing can be highly specific but is extremely labor-intensive and will probably never succeed in indexing a significant portion of the growing collection of images (Collins, 1997). Content-based systems are not yet able to identify, and therefore index, specific objects, persons, etc. and are therefore most useful for locating visually similar images. Because of these limitations, many systems rely on browsing methods to facilitate final selection.

The human vision system is capable of rapid scanning and identification of objects of interest. Potter & Levy (1969) report that the probability that an image will be retained (remembered as having been seen) depends strongly on the length of time that the image is viewed. They found sequence, presentation length of previous or following images, and total length of viewing all the images had little effect. Two seconds of viewing time resulted in almost perfect (93%) recall while 1/8 of a second yielded only a 16% "true yes" rate. The duration that the images were viewed did not have an appreciable effect on the ability of the subjects (Ss) to reject images that they had not seen (Potter & Levy, 1969). This scanning ability is exploited in browsing interfaces which present many images to the user (Hastings, 1999). Two types of browsing interface are commonly used: story board (SB) and slide show (SS). SB presentations present the images arranged in rows and columns. Scanning this type of display relies on eye movement and can result in skipping images, especially after fatigue has set in. SS presentations display the images in linear fashion (often at an unchangeable rate) with little or no opportunity to "back up" if needed.

Recent research having to do with image browsing has focused on browsing video surrogates (Cinque, Levialdi, Malizia, & Olsen, 1998) and video keyframes (Kumlodi & Marchionini, 1998); (Tse, Marchionini, Ding, Slaughter, & Kumlodi, 1998); (Lee et al., 2000); (Geisler, 2003), and improving CBIR interfaces (Jin & Kurniawati,

2001); (Stan & Sethi, 2003). The rate at which still images can be browsed and the effects of size, location, bit depth and other factors have largely been ignored. It is these questions that the experimental instrument described in this paper has been designed to investigate.

Development of the Instrument

Design Goals

The original impetus for this project came during discussion after the ASIST 2001 conference presentation by Gary Geisler (Yang, 2003). The notion that an image browser combining the SB and SS modes of presentation might be an improvement over other designs came up. In this new interface, each image would move from one display cell to the next thereby allowing several images per unit time to be displayed while each image remains on the screen for a longer period of time. I call this combined (SB and SS) mode Multiple Scrolling Thumbnails and the desire to test it necessitated the creation of a program that could do this.

The design goals of the instrument were to allow easy manipulation of the following:

- Size of the images
- Display duration
- Number of images on screen
- Bit depth

Development Environment

For lack of a better name the instrument became known as “ibi” for Image Browsing Interface. This name has stuck and will therefore be used for the remainder of this paper. Ibi was developed in AppleScript Studio™ in the Xcode environment under Mac OS X (Apple, 2003). This choice was dictated primarily by the desire to prototype rapidly without sacrificing functionality or execution speed. In addition, Apple Computer’s excellent integrated development environment, Xcode, is free. Most of the development was done on a 17” 800 MHz iMac running Mac OS X with 1024MB RAM. Some development was done on a dual USB 600MHz iBook.

Program Design and Operation

The program can be configured in two ways – through a configuration dialog or a configuration file. If the program finds a configuration file when it starts up it loads the parameters from it. If it doesn’t find the file, or if the user wants to change the configuration, a configuration (dialog) sheet (Figure 1) can be invoked. This allows all of the configurable parameters (Table 1) to be set and saved to a configuration file. This text file can then be used to configure the program in subsequent runs. The configuration file can also be manually created or edited in any program that can save plain text files.

The program window is 1004 x 720 pixels and shows a single “Start” button centered at the top when it opens. After receiving instructions the S clicks the start button with the mouse (or alternatively presses the Enter key).

This starts the display of the images. Clicking again on the button (whose name has changed to “Stop”) or pressing the Enter key stops the display and ends the test. The program then saves the data to a data file (which it creates if necessary). The data file is a tab delimited text file that can be imported into a spreadsheet or other data analysis system. The fields are shown in Table 2.

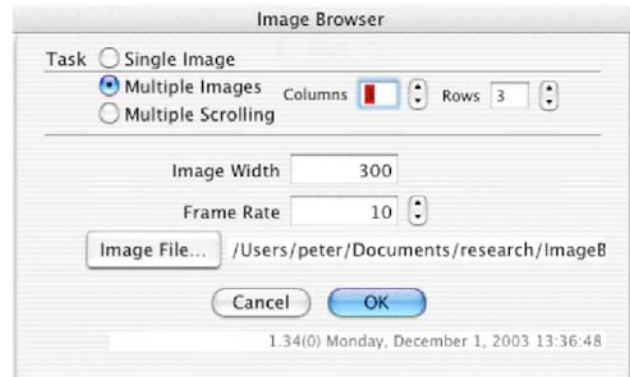


Figure 1: Ibi configuration dialog sheet

Setting the playback rate of the movie controls the rate at which the images are displayed. Since the default playback rate of each movie was set to one frame per second when the movie was built, setting the playback rate to any value other than 1 will cause the movie to display at that number of frames per second (fps). Fractional settings are allowed. Therefore a playback rate of .5 would result in .5 frames per second or 1 frame every two seconds. A playback rate of 8 results in 8 frames per second.

Table 1 – Configurable parameters

Parameter	Values
Mode ID	Described below
Number of rows ¹	1 ... 12
Number of columns ¹	1 ... 12
Number of images per second that will be displayed (fps)	0.5 ... 16
Width of the thumbnails (the aspect ratio is held constant at 4/3)	1 ... 1200
File containing the test images.	Described below

The number of images viewed was calculated by the program by multiplying the frame rate by the number of seconds between the start and stop actions. There is, of course, some potential inaccuracy introduced in this method at higher frame rates due to the program’s ability to time events only to the nearest second. This did not seem to have an effect on the results. Future versions of the program will incorporate frame-accurate timing.

¹ The total number of images, (rows times columns) is limited to a maximum of 12.

Table 2 – Data fields recorded

Item	Description
Subject Number	an integer
Date and time	e.g. Tuesday, December 2, 2003 16:26:33
Number of Columns	An integer
Number of Rows	An integer
Mode ID	Text
Completion time (in sec.)	An integer
Image width	An integer
# of images viewed	An integer

Modes of Presentation

As previously stated, this program was originally conceived to provide a means of testing three different modes of image browsing. Therefore the program was developed to display images in any of these three modes. These modes are the standard storyboard and slide show and a new mode that we call multiple scrolling thumbnails (MST). The first two (especially the storyboard) are commonly used in image browsing interfaces, (e.g. Google™, gettyimages.com, and many others). The third has not been implemented in any system to my knowledge.

The Experiment

The initial experiment was designed to test the operation and robustness of the system. The results from this experiment were also compared to results from similar work on image recognition (Potter & Levy, 1969). Twenty-six Ss performed four “systematic browsing” (Marchionini, 1995) tasks at varying frame rates.

Methodology

The Images

Images came from two sources, the *AR Faces Database* (Martinez & Benavente, 1998) and a collection of copyright-free images (a mixture of color and B&W) gathered from the web for an image indexing project (Jorgensen & Jorgensen, 2002). A sample (the *faces set*) of 306 B&W images was extracted from the faces database. This set contained images of each of 102 individuals in three standard facial expressions and lighting conditions. From these one male and one female (both smiling) face was chosen as the target images (Figure 4). The sample was randomized and compiled into a QuickTime movie of one frame per image (with a default frame rate of one frame per second.) The test movie therefore contained three different images of each face of the targets – the target image and two alternate target images. In both cases the target image appeared after the two alternate targets in the sequence. The target male face was image number 196 and its alternate targets were numbers 47 and 93. The target female face was number 186 and the alternate targets were 23 and 89.

A second QuickTime movie was created using the 904 randomized images from the web (the *general set*). It, too, was built using one frame per image and a playback rate of one frame per second. A B&W photograph of Mark Twain (in profile) was chosen as one target and a photo of a monumental arch was chosen as the second target for the general set (Figure 5). There is only one image of Mark Twain in the set and it is the 276th image. Seven images containing an arch appear before the target arch picture and 32 similar images appear after it.

1.1.1.1 Subjects

26 volunteer subjects participated in the study. The majority of them were students taking digital networking courses who received extra credit for participation. A few Ss were faculty colleagues of the researcher.



Figure 4: Target (top) and alternate target images from faces set.

The Tasks

In addition to a practice run, each S performed four identification tasks (Kwasnik, 1992): two face recognitions, and two subject recognitions. The instructions were to press the enter key (i.e. stop the system) when the S saw the exact target image. Ss were informed that similar images might appear and they should respond when they saw the exact target image. Each target image remained on the screen, outside the browser window, during the task (Figure 6). The tasks were done in the same order: first the male face, then the female face, next the photo of Twain and finally the photo of the arch. The images in the two sets were presented in the same order for each task and for all Ss. Therefore, during the second trial in each set the S was viewing the

images for the second time. The only variable that was manipulated was the number of frames per second. All images were presented at a width of 300 pixels (5.5cm). 88 trials were done in slide show mode; the remainder were done in storyboard mode. No trials were done in MST mode. The practice run consisted of a series of numbers beginning with “1” and progressing naturally (by increments of 1). The practice task was to respond when the number “25” was seen.



Figure 5: Target images from *general set*.

The equipment

All tests were performed on an 800MHz 17inch G4 iMac running Mac OS X (10.3) in the researcher’s office. The screen was set to display 1440x900 pixels at a bit depth of 16 bits per pixel and color display profile “iMac.” The width of the screen allowed space to the left of the experiment window for display of the target image. The Ss sat at a normal viewing distance (~70cm) from the display. The keyboard was on a keyboard shelf slightly below the desk height with the mouse on the righthand side next to it.



Figure 6:
Screenshot of test in progress

Results

In an effort to reduce the confounding effects of multiple variables, storyboard trials were eliminated from the data. This left 88 trials (not including the practice runs) for data analysis. There were 47 faces set trials and 41 general set trials (20 Twain & 21 arch). Due to the similarity of the faces set tasks they are grouped together without regard for the gender of the target image face. Due to the small sample size and large number of similar arch images (resulting in many in correct responses even

at the lowest frame rate) the Arches set trials will not be considered.

In the experimental trials a hit was scored when the S responded less than 3 seconds after the target image appeared. A near hit was scored when the S responded less than 3 seconds after an alternate target image appeared.

Practice Run

Based on the results from the practice run, none of the Ss had any difficulty understanding the operation of the system. Although some responded in anticipation, i.e. before the number 25 actually appeared, it was obvious that they were trying to demonstrate quick reaction times. The practice task was the only one in which the appearance of the target image could be predicted with any level of certainty.

Faces set

Four Ss recognized the target image at 1fps, 3 at 2fps and 3 at 4fps. No S responded to the target at the higher frame rates. Near hits (responses to alternate target images) were recorded for all but the highest frame rate. The results for the faces set trials for each frame rate are summarized in table 3. The responses to the two alternate target images are combined in the near hits count.

Twain set

Two Ss or 100% of those viewing the images at 1fps responded correctly. 3 out of 7 Ss correctly responded at 2fps, 3 out of 10 responded correctly at 4fps and none at 8 or 12fps. No subject performed this task at 16fps. The results for the Mark Twain set for each frame rate are summarized in table 4.

Table 3 – Results for Faces trials, N is number of Ss, No. is number of correct responses, % is percent of all Ss at that frame rate who responded correctly.

Fps	N	Hits		Near Hits		Total	
		No.	%	No.	%	No.	%
1	7	4	57%	2	29%	6	86%
2	10	3	30%	2	20%	5	50%
4	14	3	21%	3	21%	6	43%
8	7	0	0%	1	14%	1	14%
12	6	0	0%	1	17%	1	17%
16	3	0	0%	0	0%	0	0%
Total		47					

Discussion

In pilot testing it was discovered that expecting the Ss to click on a button to indicate their response (i.e. stop the display as they had seen the target image) was problematical as they often inadvertently moved the mouse without knowing it after starting the trial. Therefore the program was altered so that pressing the

enter key on the numeric keypad (which is at the extreme righthand end of the keyboard) would also start and stop the trial. All Ss were instructed to use the enter key to control the program.

Table 4 – Results for Twain trials, N is number of Ss, No. is number of correct responses, % is percent of all Ss at that frame rate who responded correctly.

Fps	N	Hits	
		No	%
1	2	2	100%
2	7	3	43%
4	10	3	30%
8	4	0	0%
12	2	0	0%
16	0	NA	NA
Total	25		

Although the think-aloud protocol was not used many Ss involuntarily made some exclamation (such as “Oh Golly!”) when a high rate (8, 12 or 16fps) display started. Several made comments such as “No way!” or “holy cow!” These comments combined with the low success rates at higher frame rates as measured in this experiment lend considerable support to Potter’s (Potter & Levy, 1969) results regarding 8fps being a practical maximum. While a larger sample would provide a finer measure of the limit it is safe to say that frame rates of 12fps and over (and probably down to 8fps) are too fast for the average person performing a systematic browsing task in slide show mode with 5.5cm wide images at a viewing distance of 70cm.

The data, when aggregated, support Potter’s finding that under these conditions the recognition rate decreases linearly with increasing frame rate (Figure 7).

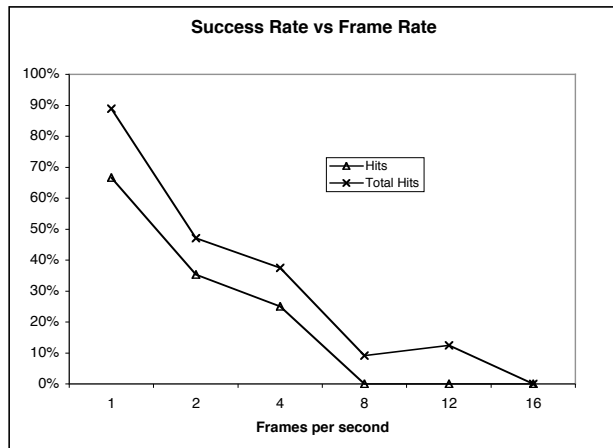


Figure 7: Hit rate and Total Hit rate vs Frame Rate

Conclusion

A flexible test instrument has been developed for testing image-browsing performance under a variety of conditions. The instrument is easy to use and highly configurable. Preliminary tests with the instrument have yielded data which is in keeping with earlier research in human perception. Future research in the effects of image size, display mode, bit depth and other presentation variables on browsing performance can be now carried out in a standardized way. This researcher has begun testing the effect of image size and number of thumbnails presented at once. Multiple scrolling thumbnails will also be tested by this researcher in future experiments. An important addition to this methodology would be a large, diverse, thoroughly indexed and annotated image database standard.

ACKNOWLEDGMENTS

The project was supported in part by the Florida State University Center for Research and Creativity First Year Assistant Professors Program and the Information Use, Management & Policy Institute of the School of Information Studies at Florida State University.

REFERENCES

Apple Computer. (2003). AppleScript Studio (Version 1.3) [programming language]. Cupertino, CA: Apple Computer.

Berinstein, P. (1999). Do You see What I see? Image Indexing Principles for the Rest of Us. Online, 1999(March-April), 85-88.

Cinque, L., Levialdi, S., Malizia, A., & Olsen, K. A. (1998). A Multidimensional Image Browser. *Journal of Visual Languages and Computing*, 9, 103-117.

Collins, K. (1997). Providing Subject Access to Images: A Study of Users Queries. Unpublished Masters, University of North Carolina, Chapel Hill, NC.

Geisler, G. (2003). AgileViews: A Framework for Creating More Effective Information Seeking Interfaces. Unpublished PhD, University of North Carolina, Chapel Hill, NC.

Google. (2004, 2004). Google Image Search. Retrieved Jan 29, 2004, from <http://images.google.com/>

Hastings, S. K. (1999). Evaluation of Image Retrieval Systems: Role of User Feedback. *Library Trends*, 48(2), 438-452.

Jin, J. S., & Kurniawati, R. (2001). Using Browsing to Improve Content-Based Image Retrieval. *Journal of Visual Communication and Image Representation*, 12, 123-135.

Jørgensen, C., & Jørgensen, P. (2002). Testing a Vocabulary for Image Indexing and Ground Truthing. Paper presented at the Internet Imaging III, San Jose, CA.

Komlodi, A., & Marchionini, G. (1998, June 23 - 26). Key fram preview techniques for video browsing. Paper presented at the International Conference on Digital Libraries, Pittsburgh, PA.

Kwasnik, B. H. (1992, 10-14 August 1992). A Descriptive Study of the Functional Components of Browsing. Paper presented at the Proceedings of the IFIP TC2/WG2.7 Working Conference

- on engineering of Human-Computer Interaction, Ellivuori, Finland.
- Lee, H., Smeaton, A. F., Berrut, C., Murphy, N., Marlow, S., & O'Connor, N. e. (2000). Implementation and Analysis of Several Keyframe-Based Browsing Interfaces to Digital Video.
- Marchionini, G. (1995). Information Seeking in Electronic Environments. Oxford: Oxford University Press.
- Martinez, A. M., & Benavente, R. (1998). The AR Face Database: CVC Technical Report #24 (Technical).
- Potter, M. C., & Levy, E. I. (1969). Recognition Memory for a Rapid Sequence of Pictures. *Journal of Experimental Psychology*, 81(1), 10-15.
- Stan, D., & Sethi, I. K. (2003). eID: a system for exploration of image databases. *Information Processing and Management*, 39, 335-361.
- Tse, T., Marchionini, G., Ding, W., Slaughter, L., & Komlodi, A. (1998). Dynamic key frame presentation techniques for augmenting video browsing. Paper presented at the working conference on Advanced visual interfaces, L'Aquila, Italy.
- Yang, M., Wildemuth, B. M., Marchionini, G., Wilkens, T., Geisler, G., Hughes, A., Gruss, R., and Webster, C. (2003, Oct 19-22, 2003). Measuring User Performance During Interactions with Digital Video Collections. Paper presented at the 66th Annual Meeting of the American Society for Information Science and Technology (ASIST 2003), Long Beach, CA.