Carnegie Mellon University
Research Showcase

Human-Computer Interaction Institute | School of Computer Science

1-1-2002

Interfaces for Palmtop Image Search

Mark Derthick
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/hcii

Recommended Citation
Derthick, Mark, "Interfaces for Palmtop Image Search" (2002). *Human-Computer Interaction Institute.* Paper 27.
http://repository.cmu.edu/hcii/27

This Conference Proceeding is brought to you for free and open access by the School of Computer Science at Research Showcase. It has been accepted for inclusion in Human-Computer Interaction Institute by an authorized administrator of Research Showcase. For more information, please contact research-showcase@andrew.cmu.edu.

Proceedings of the Joint ACM/IEEE Conference on Digital Libraries, Portland, OR, July, 2002, pp. 340-341.

Interfaces for Palmtop Image Search

Mark Derthick
Human Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA, USA 15213
1 412 268-8812
mad@cs.cmu.edu

ABSTRACT

Will current technology support search for video news or entertainment on mobile platforms? An Ipaq palmtop version of the Informedia Digital Video Library interface has already been developed at the Chinese University of Hong Kong. For these displays, the desktop technique of showing a large grid of images in parallel is not feasible. Perceptual psychology experiments suggest that time-multiplexing may be as effective as space-multiplexing for this kind of primed recognition task. In fact, it has been specifically suggested that image retrieval interfaces using Rapid Serial Visual Presentation (RSVP) may perform significantly better than parallel presentation even on a desktop computer [2]. In our experiments, we did not find this to be true. An important difference between previous RSVP experiments and our own is that image search engines rank retrievals, and correct answers are more likely to occur early in the list of results. Thus we found that scrolling (and low RSVP presentation rates) led to better recognition of answers that occur early, but worse for answers that occur far down the list. This split confounded the global effects that we hypothesized, yet in itself is an important consideration for future interface designs, which must adapt as search technology improves.

Categories and Subject Descriptors

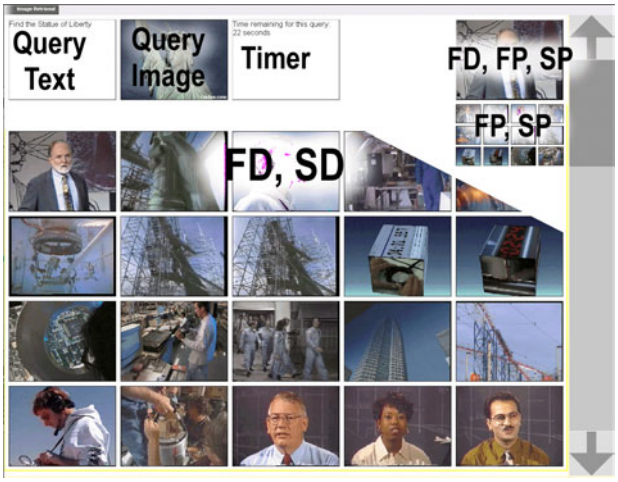
H.5.2 [**User Interfaces**]: Graphical user interfaces (GUI), Interaction styles.

General Terms

Algorithms, Human Factors, Image Retrieval

1. Previous Work

Although our goal is video retrieval, our current interfaces segment the video into shots, and represent them with single frames. Therefore the present work addresses image retrieval exclusively. We expect the user to have a query in mind, and that the mental representation of the query will prime the recognition of matching images. Studies have shown effective recognition at image presentation rates of 63ms [2]. In contrast, making an eye movement and fixating on an adjacent image is expected to take about 250ms [2]. Previous experiments with sequential presentation of video keyframes have looked at summarizing the content of a video [1]. This is significantly different from our task both because there is no priming involved, and because the images are from contiguous shots in a single video. In any case, these experiments have not found that either serial or parallel presentation dominates unconditionally.



2. Interface Design

Although our primary interest is in palmtop interfaces, we wanted to find out how much worse palmtop performance would be compared to desktop performance. We hypothesized that there would be an interaction; that RSVP compared to parallel presentation would be relatively better on a palmtop than a desktop. This is because in RSVP the image size is the same on palmtop and desktop, while the scrolling images on the palmtop are much smaller.

We designed for an Ipaq, which has 240 x 320 pixels, each 0.24mm square. In order to focus on the serial vs. parallel perception issue, and isolate it from issues of controlling the interface, as well as to make the 4 interface layouts as similar as possible for the subjects, we simulated the Ipaq layout on part of a 21" desktop screen at 1600x1200 resolution. We used a very large scrollbar in the same location for all interfaces, and we used the same placement and layout to present the query text, query image, and countdown timer. Figure 1 shows the union of the 4 layouts, abbreviated SD (scrolling desktop), FD (flashing desktop), SP (scrolling palmtop) and FP. The large image grid was continuously visible for SD, and only visible when the RSVP slideshow was stopped in FD. The large image in the upper right

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL'02, July 13-17, 2002, Portland, Oregon, USA.

Copyright 2002 ACM 1-58113-513-0/02/0007…$5.00.

was always visible in the FD, FP, and SP versions. Clicking on it started or stopped the slideshow for FD and FP, and had no effect for SP. Clicking on a grid image toggled a yellow highlight around the image, which indicates selected images. Because there is a delay in stopping the slideshow after recognizing an image, we show the last 12 images seen in the grid. For FP this required using 3 rows of 4 images, each only 56x42 pixels. We used the same size for SP, and also retained the large picture. In both FP and SP, clicking on a small image selected that image into the large picture, so that a better correctness judgment could be made. Twenty subjects participated in a within-subject randomized design. They watched a 4 minute movie explaining the interface and task, carried out 4 practice questions (1 with each layout), 15 analyzed questions, and completed pre- and post-test questionnaires. Up to 200 relevance ranked images had previously been found with an automatic image similarity algorithm for each query image drawn from the TREC-2001 Video Retrieval Track, of which between one and fourteen had been hand-labeled as correct. Subjects had 30 seconds to scroll or flash through the images and click on those they thought relevant. The goal was to have more images than we expected could be searched so that we would measure the kind of throughput we had found necessary in the interactive TREC competition. Score was average F-measure over the 15 questions. There was a huge variation in question difficulty. For some, there was only a single correct answer buried far down in the list. For others the top five images were correct and there were no other correct answers. Subjects spent about half an hour each. The subject with the highest score was given $100, and the others received no compensation.

3. Experimental Results

Considering only the scrolling versions, the desktop interface was significantly better than the palmtop interface (means 0.53 vs. 0.42, $p < 0.01$). Considering only the flashing versions, the 200ms rate was marginally better than the 140ms rate (means 0.47 vs. 0.42, $p < 0.06$). No other single-factor differences were significant, and neither was the hypothesized interaction between desktop/palmtop and scrolling/flashing. However by looking at finer-grained variables we developed an explanation for the lack of significance. Some interfaces allowed more time to look carefully at the top results. Because there were more correct answers near the top, these interfaces performed better overall. However for finding images far down the list, the results are different. For each of the 91 correct answers spread across the 15 questions, we plotted the number of subjects who found the answer and grouped by the interface they used. For all the interfaces, there is an approximately linear decrease in the odds of finding an answer as the answer is buried farther into the result list. Figure 1 shows the regression lines. SD has the highest intercept and the slowest drop-off, and therefore dominates all the other interfaces. FD200 and FP200 have almost identical parameters, having the second best intercept, but the worst slope. This makes sense; a slower slideshow gives you more time to pick accurately for the images you do see, but you don't see so many images. SP and FD140 are almost identical, with the next best intercept, and better slope. Worst overall is FP140, with the lowest intercept, but the same slope as SP and FD140.

Most subjects hated the flashing interfaces, especially for the palmtop. However a minority loved the flashing interfaces, especially for the palmtop. Subjects with experience working with images scored higher and were more likely to prefer the flashing interfaces. This suggests follow-up experiments that include extensive training. Learning effects were noticeable, with an average improvement of 0.09 over the 15 questions.

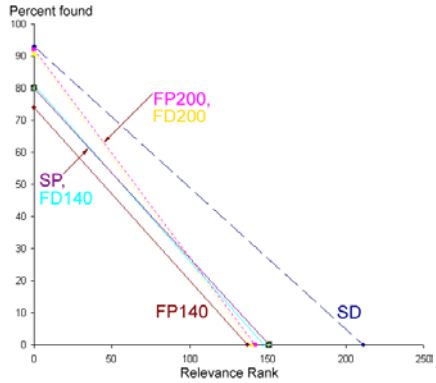


Figure 2 Good performance on high rank results doesn't hold up for lower rank results for 200ms flash rates.

4. Conclusion

Characterizing performance in terms of slope/intercept for trading off good performance on the first few results against reasonable performance on long result lists suggests that interfaces need to be tuned to match the underlying automated search algorithm. It is encouraging that overall performance on a palmtop is in the same ballpark with desktop performance (means 0.47 vs. 0.42). For experimental purposes we forced users to either flash at a pre-determined rate or scroll. If flashing is controllable like an automobile's cruise control it offers advantage of rate-consistency when there is no interesting "traffic" with the option of fine manual control when there is. The ability to scroll backward reduces the need to show 12 images in the grid, and larger images in a 3x2 grid may increase palmtop performance. It seems clear that no scrolling palmtop interface can perform as well as a scrolling desktop interface. Some version of serial presentation seems like the best hope for effective palmtop interfaces. In this respect, subjects' subjective impressions are discouraging. We expect that much more experimentation will be carried out on palmtop video retrieval. These results suggest that testing the best overall design in order to address user acceptance may be more important for the near term than the factorial designs we used to compare performance across designs.

5. Acknowledgements

This material is based on work supported by the National Science Foundation under Cooperative Agreement No. IRI-9817496. Thanks to the Informedia research group and the anonymous reviewers for improvements to the presentation.

6. References

1. Christel, M.G., A.G. Hauptmann, A.S. Warmack, and S.A. Crosby. *Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library*. in *Advances in Digital Libraries*. 1999. Baltimore, MD: IEEE: p. 98-104.

2. Ware, C., *Information Visualization: Perception for Design*. 2000: Morgan Kaufmann.