

Visual Interface and Control Modality: An Experiment about Fast Photo Browsing on Mobile Devices

QianYing Wang, Susumu Harada, Tony Hsieh, Andreas Paepcke

Stanford University
Stanford, CA 94305
United States of America
wangqy@stanford.edu, {harada, thsieh, paepcke}@cs.stanford.edu

Abstract. We examined the strengths and weaknesses of three diverse scroll control modalities for photo browsing on personal digital assistants (PDAs). This exploration covered nine alternatives in a design space that consisted of three visual interfaces and three control modalities. The three interfaces were a traditional thumbnail layout, a layout that placed a single picture on the screen at a time, and a hybrid that placed one large photo in the center of the display, while also displaying a row of neighboring thumbnails at the top and bottom of the screen. In a user experiment we paired each of these interfaces with each of the following three scroll control modalities: a jog dial, a squeeze sensor, and an on-screen control that was activated by tapping with a stylus. We offer a simple model that classifies our experiment's interfaces by how much they provide visual context within the photo collection. The model also classifies the scroll modalities by how tightly they correlate scroll input actions to effects on the screen. Performance and attitudinal results from the user experiment are presented and discussed.

1 Introduction

As digital cameras become increasingly prevalent, large personal libraries of digital photographs are becoming more and more common. The low incremental cost of digital photography tempts photographers into accumulating photographs faster than they ever did before. This growth in digital photograph libraries has pushed interested parties to seek new ways to store, show, and retrieve digital images. As personal digital assistants (PDAs) with credible processing, storage, connectivity and display capabilities emerge, these devices are becoming potential platforms that enable users to have their entire digital photo collection available to them at all times.

The potential for PDA-based photo collections reaches beyond personal photographs. Environmental biologists, anthropologists, real-estate agents, and work crews that document the city-wide condition of telephone poles are all examples of mobile workers who rely on photographic evidence that needs to be accessible in the field.

There are, however, still a number of questions whose answers will lead to improved designs of small-display photo browsers. An appropriate approach needs to take into account two very different, but interacting aspects: the *visual interface* and *physical control modality*. The visual interface refers to the presentation of photos to the human viewer with the goal of facilitating browsing, conducting a focused search, or studying a picture in detail. Control modality refers to the physical mechanism that allows viewers to communicate their intentions to the device. Examples of such intentions are “scroll through this deck of photos, or “make this photo bigger.”

The most widely used photo interface on PDAs and desktops is a grid of thumbnails that is displayed on the screen. A number of alternative interfaces have been developed for browsing photos on desktops, such as an approach that lets viewers zoom in and out of images [1]. Most desktop interfaces, however, do not perform well when simply replicated on a PDA without modification.

Two of the common control modalities for PDAs are the stylus, which is used to tap or drag items on the screen, and the jog dial. The latter is a small wheel that protrudes from the PDA housing in a position where the user's thumb can reach (see Figure 1-A). The thumb rolls the dial up or down. Some wheels can be rolled indefinitely; others are spring-loaded and have three operating positions: up, down, and neutral. They are sometimes also referred to as *jog wheels* and *scroll dials*, or *scroll wheels*.

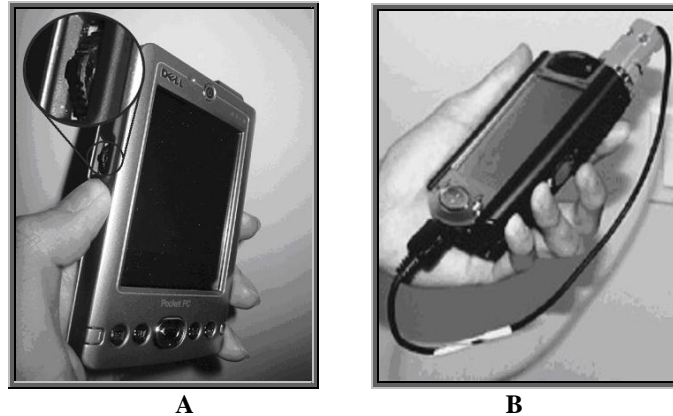


Fig. 1-A: Jog Dial Input Device, **1-B:** PDA with squeezable input device (circular sensor below the middle finger)

We constructed a different control modality that allows its operator to squeeze the PDA with (usually) the index or middle finger while the device rests in the palm of the hand (Figure 1-B) [2]. The exerted pressure is recorded and transmitted to the PDA software. Applications can use this continuous pressure data to control, for example, the scrolling speed, the frame rate of animations, or the zoom factor of a graphical user interface window. Please see Related Work Section for more references to control modalities.

In an effort to understand the strengths of several diverse interfaces and control modalities, as well as the interactions among them, we implemented three PDA photo browser interfaces and constructed or acquired PDAs with three different control modalities. We then conducted an experiment in which participants were asked to find photos using each combination of interface and control modality.

In the following sections we describe our choices for interfaces and modalities. We then explain the theory that helped us decide on the interfaces and modalities to deploy for our experiment, and produced the predictions that we carried forward into the investigation. Finally, we lay out our experimental design, describe our measured results, offer a discussion of findings, and conclude.

2 Interfaces and Modalities

We finalized on three modality types, to be compared within the context of three interface types. This arrangement resulted in nine applications, each a unique combination of a particular modality and a particular interface, represented as two-letter abbreviations in Table 1. Table 1 lists the three modality types along the top-most row, and the three interface types along the left-most column. We present here the descriptions of each of the interface types, followed by the description of each modality type.

Table 1. Three modality types and three interface types, and the resulting nine applications (experimental conditions)

Modality Interface	Squeeze (S)	Jog (J)	Click (C)
Thumbnail (T)	TS	TJ	TC
Parade (P)	PS	PJ	PC
Fullscreen (F)	FS	FJ	FC

2.1 Interfaces

The first interface is called the **Thumbnail** interface, in which the thumbnails of the photos are arranged in a grid with a fixed number of rows and columns (Figure 2-A). This interface is, on most existing photo

browser applications, the primary presentation style. One difference here is that we replaced the typical scroll bar, which normally allows the user to scroll up and down a multi-page collection, with what we call a context bar. A context bar is a user interface component that is similar to the scroll bar, except for the fact that it does not accept any user input and only serves to provide contextual information about the current position of the corresponding viewing area with respect to the entire length of the content area. We made this change not in advocacy of the context bar as a user interface facility, but to allow proper control over the independent variables of our experiment. The photos are ordered left to right, top to bottom.

The second interface is called the **Parade** interface (Figure 2-B). In this interface there is one photo at the center of the screen that is larger than the rest of the thumbnails. Above and below this central photo, we place one row of smaller thumbnails. The photos in this interface are also arranged left to right and top to bottom. As the user navigates through the photo collection, the photos "parade" along in either direction, following a trajectory of the letter "Z" with the top and bottom horizontal part of the letter corresponding to the top and bottom rows of thumbnails (Figure 2-C). During the diagonal part of the trajectory, the photo enlarges or shrinks, depending on whether it is approaching or moving away from the center of the screen.

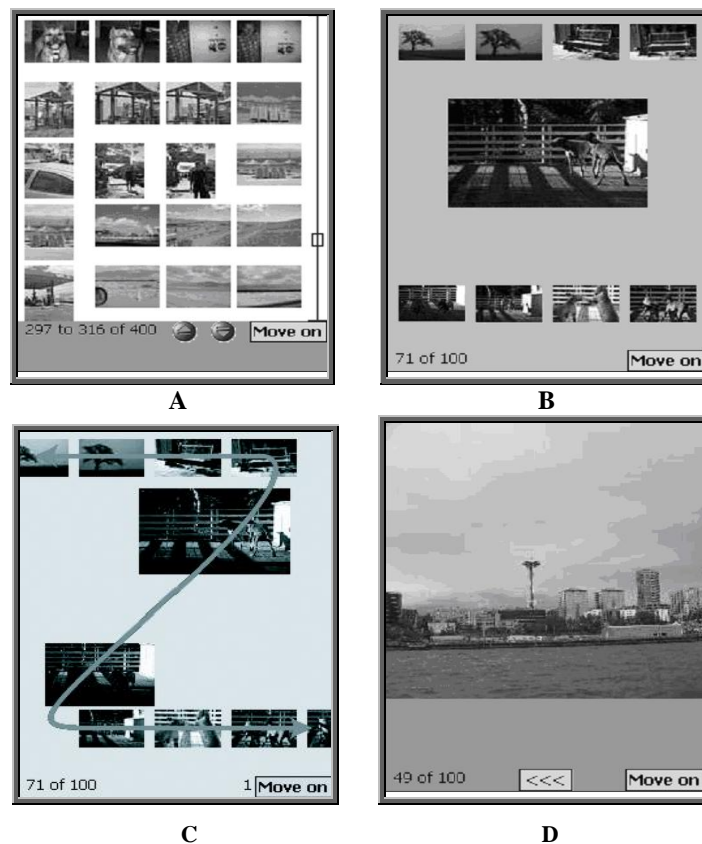


Fig. 2-A: Thumbnail interface. **2-B:** Parade interface. **2-C:** Scrolling animation of Parade interface. Photos animate along the "Z" shaped motion path. The drawn-in path in the figure is not visible on the actual interface. **2-D:** Fullscreen interface

The third interface is the **Fullscreen** interface (Figure 2-D), which simply displays each photo at the maximum size that fits within the PDA screen. From the user's perspective, it is as if the user is looking at one particular photo at a time along a horizontal filmstrip, where the photos are ordered from left to right. As the user navigates through the photo collection, the filmstrip scrolls towards the left or right.

For each of the three interfaces presented above, we incorporated the three modalities, described in detail below.

2.2 Modalities

The first modality is the **Click** modality, which represents the prevalent modality for flipping through photographs on existing PDA photo browsing applications. The key aspect of this modality is that the user has to explicitly click a "next" or "previous" button repeatedly in order to scroll through the photo collection. Under certain scenarios, the user may be able to hold down on the "next" or "previous" button in order to continuously scroll in one direction, as described later in this section. Under this modality, the screen shows a pair of left/right arrows (up/down arrows in the case of the Thumbnail interface), which the user can click, using the stylus. One click on either one of the arrows results in the photo collection being shifted by one "increment" in the corresponding direction. One increment corresponds to one photo, except in the case of the Thumbnail interface where it corresponds to one row of photos. The Thumbnail interface is one example where the user can tap and hold one of the arrows, resulting in continuous scrolling, mimicking the behavior of typical scrollbar arrows in most thumbnail based photo applications. For the other interfaces, the user clicks on the arrow once for each increment of change. We felt that this was a fair reflection of the existing photo applications, especially in the Fullscreen interface, as most applications that offer such "slideshow" capabilities only allow the user to move to the next or previous photos one at a time.

The second modality is the **Squeeze** modality, using the squeeze input device we developed [2]. In this modality, although the manipulation occurs through the external squeezable input device, the mapping between the user's squeeze pressure and the effect on the photos on the screen is unidirectional. That is, as the user squeezes harder, the photos scroll by faster, and if the user releases the squeeze, the photos stop scrolling, but will not change the direction of the scroll. We placed a direction indicator button at the bottom of the screen, which the user can tap with the stylus to toggle the direction of the scroll. Whenever the user taps this button, the direction toward which the photos will scroll changes and the label on the button reflects this new direction as well. Generally, subjects held the PDA with their left hand, as shown in Figure 1-B, using the middle or index finger to control the squeezable input device.

The third modality is the **Jog** modality, in which a jog dial is used to navigate forward or backward through the photo collection. The Jog modality is the closest "competitor" that exists in the market today to the Squeeze modality, with the key characteristic of being able to be operated via the hand that holds the PDA. One marked difference between the Squeeze modality and the Jog modality is that the Squeeze modality allows for a continuous range of input values, depending directly on the squeezing pressure applied by the user, whereas the Jog modality only provides a sequence of up/down events at a constant rate when the dial is held up or down, away from its resting position.

3 Hypotheses

We partition the areas of our concern into two main categories: *Cognition* and *Manipulation*. Cognition refers to characteristics of the *interfaces* that affect the user's ability to comprehend what is on the display, and to apply the information towards the goal of finding a given photograph. By Manipulation we mean aspects of the *control modalities* that affect the user's ability to manipulate the interface effectively.

3.1 Cognition

Both the Cognition and Manipulation categories comprise numerous facets, which are under study by other disciplines of inquiry. For our purposes we concentrate on two aspects of the Cognition category that are particularly important in the context of image search: the user's *sense of place* and the degree of *attention focus* that an interface elicits.

Sense of place refers to the ability of a user to know which portion of the overall collection is being displayed on the screen at any given time. This might be the understanding that the visible images are part of a particular birthday party, or cover some particular time frame. One common method for increasing sense of place is to provide context. For example, fish-eye techniques [3] provide the user with visual clues for what is near a displayed information item of interest.

The degree of **attention focus** is the amount of attentional resource a viewer can allocate to each information unit to absorb in more detail. For example, the thumbnail view requires a broad sweep before

the onlooker can pick a photo to examine more closely (low attention focus). In contrast, a single photo on the screen allows for high (sharp) focus.

Figure 3 places the three interfaces of the previous section into this Cognition space. The ordinate tracks sense of place, the abscissa marks attention focus. The values of the two variables are discrete; they can be thought of simply as 'low,' 'medium,' and 'high.'

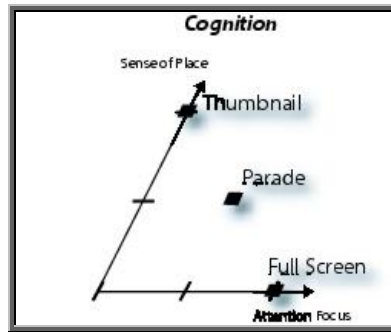


Fig. 3: Cognition: Sense of Place vs. Attention Focus

The figure shows why we chose the three interfaces we introduced earlier. They represent two extremes and one 'compromise' solution in the Cognition space. The thumbnail interface, as pointed out earlier, provides significant sense of place, but requires the viewer to split his attention across the entire display. Fullscreen allows the viewer to examine the (only visible) photo almost immediately, allowing high attention focus at the expense of sense of place. Parade, with its large photo in the center and smaller thumbnails above and below attempts a middle ground for both quantities.

3.2 Manipulation

The quantities we highlight for the control manipulation category are *transport speed* and *control-effect coupling*. **Transport speed** is the speed with which a given control modality allows the user to move between photos. Figure 4, analogous to Figure 3, places the three control modalities we chose into the Manipulation space. Squeeze enables the viewer to advance a sequence of images at various desired speeds. Pressure controls transport speed on a continuous scale. Transport speed with Click, in contrast, is limited by the operator's rapid tapping ability. While considerable, this speed is no match for the transport blur that a computer can generate. Our compromise between these extremes is the Jog control manipulation. Its on/off nature requires the computer to limit the transport speed such that the 'average' user can follow. Users are therefore not able to accelerate to blurring speed when they know that they are far from the target photo, and to slow down when they reach the image's neighborhood.

Control-effect coupling indicates how closely a discrete manipulative action matches a consequent identifiable event on the display. Click, for example, has high control-effect coupling: one tap with the stylus advances photos by one unit, for instance a single photo in Fullscreen, or one row of thumbnails in the Thumbnail interface. The Squeeze modality has the loosest coupling of the three modalities. Pressure on the sensor controls transport speed, which is a more indirect control than the photo position controlled with Click¹. We place Jog midway between the extremes because the jog dial's two off-center positions produce a predictable, single-speed transport. It is less coupled than Click because when you hold the jog dial in the up or down position, the transport continues without any further action on the user's part. Jog is more coupled than Squeeze because of its on/off nature.

¹ One can think of this relationship as Squeeze impacting the first derivative of the first-order effect, which is what Click controls directly.

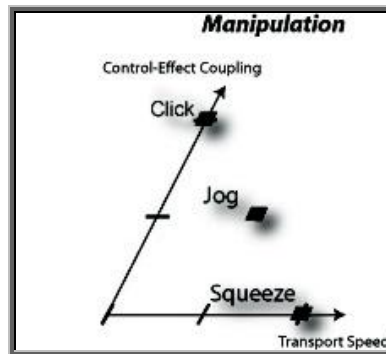


Fig. 4: Manipulation: Control-Effect Coupling vs. Transport Speed

3.2 Predictions

For both Cognition and Manipulation, a high value along either axis in their respective figures is 'good.' In both cases, however, one goal is favored only at the expense of the other: the luxury and absorption speed of high attention focus usually requires sacrifices in sense of place. Similarly, high transport speeds can lead to a lesser sense of control. Our experiment tries to illuminate where good tradeoffs are to be made when image search is the task.

Interactions between Cognition and Manipulation are equally important to understand. For example, the Thumbnail interface can likely make do with lower transport speed, because of its low attention focus and consequent absorption delays. A high attention focus interface is similarly likely to call for high transport speed controls.

Our predictions for the experiment were therefore:

1. The high sense of place Thumbnail interface will thrive with the low speed Click modality.
2. Squeeze will do well with any interface, because it allows users to control speed continuously. If the user wants to go slow for a certain interface, he can. If the user prefers high transport speed with another interface, Squeeze can accommodate him as well.
3. Squeeze will work best with the Fullscreen interface because with a high transport speed the user will need to get a quick understanding of what he is seeing. The only way this can happen is with a high attention focus.

4 Experiment

Photo browsers need to support a number of user activities, ranging from idle browsing, to searching for a particular photograph. We chose search as the task for our experiment, because it demands from the user a large number of interactions with the device. In the interest of avoiding confounding factors, we decided to enable in our experiment only facilities for linear visual search, rather than the kind of sophisticated support that we provided in [4].

4.1 Participants

We recruited 23 participants for our experiment, ranging from ages 17 to 38, with no special criteria for selection. Of the 23 participants, 17 were male and six were female, with two participants being left handed and 11 without any prior experience with PDAs.

4.2 Settings

We used two PDAs for our experiment: a Hewlett-Packard iPaq H5500 and a Dell Axim X3. Both PDAs housed a 400MHz Intel XScale processor running the Microsoft Windows Mobile 2003 operating system, with a 240x320 16 bit color display. The Axim came equipped with the jog dial (Figure 1-A), and the iPaq was adorned with our custom-built squeeze input device (Figure 1-B).

4.3 Procedures

Our experiment followed a within-participant design. We exposed each participant to nine experimental conditions, resulting from the two factors (interface and modality) with three levels each, as shown in the matrix in Table 1.

For each condition, the participant performed four search task trials, for a total of 36 trials. At the beginning of each search task trial, the participant was shown a target photo displayed full-screen on a separate PDA. We consistently used the same 1,800 image collection throughout the experiment. The photos were unknown to all participants and were divided among conditions. While using each participant's own collection would have approximated real-life situations better, this collection uniformity was necessary to control for differing familiarity of participants with their own collection. We provided the participants with a sample trial at the introduction of each new condition, where they were given the opportunity to experiment with the particular combination of interface and modality (using separate images from that of the 1,800 images in the collection).

Within a particular condition, the participant searched for four different photos, each time starting at some random point along the collection. In order to provide the participants with a sense of where they were within the photo collection at any given time, the application also displayed the index of the currently centered photo with respect to the total collection size at the lower left corner of the screen. Within each interface, we rotated the sequence in which we exposed participants to the three modalities.

4.4 Measures

We explored two categories of experimental measures: performance and attitudinal. There were two performance measures. The first measure was the average time that participants used to find a target photo under each experimental condition (modality/interface). We did not place any constraints on the amount of time the participants could spend on searching for a photo. The second measure is the success rate. Participants might have difficulty to locate a specific target photo. They could skip any trial by tapping a button labelled "move on". If participants tapped the "move on" button, this trial would be recorded as a failed trial, but the search time would not be included in the calculation of the average search time.

Participants filled out a questionnaire after the trials of each condition to indicate their subjective evaluations of the nine modality/interface combinations. The rankings for these attitudinal aspects included perceived efficiency, reliability, enjoyment, physical strain, photo size, and screen layout aesthetics.

5 Results

5.1 Performance

Figure 5 displays each modality's photo search time in seconds, averaged over the three interfaces. The average time to use the Squeeze modality to find a photo was 56.68 seconds. For Jog, the average search time was 52.16 seconds, for Click it was 58.92 seconds. Participants spent significantly less time searching for a photo with the Jog modality than with the Click modality ($t(22)=2.34, p<.05$). There was no statistically significant difference between the search time of Jog and Squeeze ($t(22)=1.76, p>.01$), and between Squeeze and Click ($t(22)=0.67, p>.01$). The success rates for the three modalities were all above 98%. No statistical differences were found among the success rates across modalities ($F(2,44)=0.87, p>.1$).

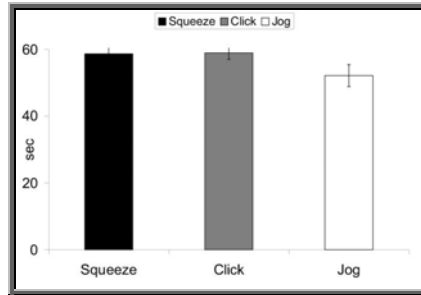


Fig. 5: Avg Search Time Across Modalities

For the Parade interface, the Squeeze and Jog modalities yielded no statistically significant differences; both consumed an average of 48 seconds. Click, on the other hand, at 62 seconds, was slower than Squeeze and Jog by about 23%. Repeated measure ANOVA showed that the difference was highly significant ($F(2,44)=9.14, p<.01$). The success rates for Parade were at a very high 98% for all modalities. That is, modality had no significant impact on how many participants were able to find the target photos across trials ($F(2,44)=0.324, p>.1$).

The Thumbnail interface was unaffected by modality for both time and success rate. Search time differences were not statistically different ($F(2,44)=0.621, p>.1$) and emerged at an average of 75 seconds, no matter whether the participants used Squeeze, Jog, or Click to interact with the photos. The same observation is true for the success rate $F(2,44)=1.0, p>.1$). As under the Parade conditions, the rate held steady at 98%.

For the Fullscreen interface, modality did have an impact. This impact differed, however, from the Parade case. When interacting with photos under the Fullscreen condition, there was no significant difference between the Squeeze and Click modalities ($t(22)=0.83, p>.01$). Their search times were both at 46 seconds. Jog, in contrast, had participants finding their photos in about 38 seconds. This search speed difference across modalities was highly significant ($F(2,44)=5.82, p<.01$). Success rates, again, were not influenced by which input modality the participants worked with ($F(2,44)=1.0, p>.1$).

5.2 Attitudinal

Our questionnaire contained a number of questions that we later collapsed into six core indices by means of a factor analysis. Four of the six indices pertained to modality type. They were Perceived Efficiency, Reliability, Enjoyment, and Physical Strain. Photo Size and Screen Layout Aesthetics were the two other indices pertaining to interfaces. All results that are reported as significant are at $p < 0.05$.

Perceived efficiency and enjoyment had similar results across all three modalities. There was no significant difference between Jog and Squeeze for perceived efficiency and enjoyment. Both modalities were rated significantly more efficient and enjoyable to use than the Click modality. Both reliability and strain are inextricably related to modality manipulation. We would expect, for example that strain would likely be caused by the pushing of the squeeze sensor. We therefore examined participants' impressions of reliability and strain separately for each modality. Click was rated as the most strenuous modality to use, followed by Squeeze and Jog. Squeeze was rated as the least reliable modality, followed by Click and Jog. The strain and reliability differences between any two modalities were statistically significant.

Results for interfaces were in agreement with the patterns for individual modalities for both efficiency and enjoyment. That is, for all three interfaces, there was no significant difference between Jog and Squeeze for perceived efficiency and enjoyment. They both were rated significantly more efficient and enjoyable than Click. For Parade and Thumbnail, the perceived efficiency ranking was consistent with the search time performance for the three modalities. A discrepancy between perceived efficiency and search time was found for the Fullscreen interface where participants rated Squeeze more efficient to use than Click, while no search time performance difference was found between these two modalities.

Reliability and strain results for the three interfaces are illustrated by Figure 6-A and Figure 6-B. In these two figures, if two bars are at the same height, there is no significant difference between the two

corresponding modalities. In other words, any height difference represents a statistically significant difference for that measure.

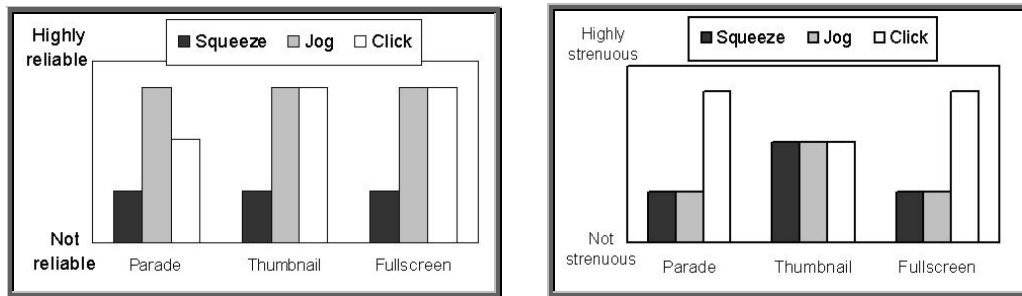


Fig. 6-A: Relative Ranking of Perceived Reliability; **6-B:** Strain Rankings

The two cross-interface measures we compared were screen layout aesthetics and photo size, (see Figure 7). The Fullscreen interface was rated the best for layout aesthetics, followed by Parade and then Thumbnail. The difference between any two interfaces was significant. Fullscreen and Parade were rated as having a more appropriate photo size than the Thumbnail interface. We measured no significant difference between the Fullscreen and Parade interfaces for photo size.

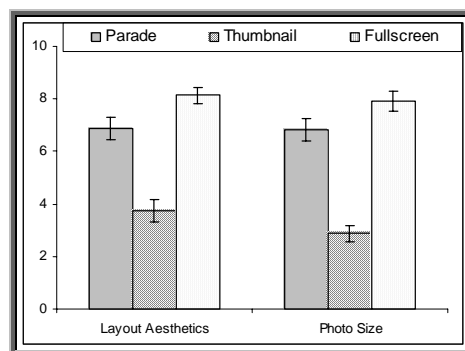


Fig. 7: Subjective Impressions of (a) layout aesthetics and (b) appropriateness of photo size

6 Discussion

6.1 Search Time and Efficiency Perception

Recall that the participants' perceived efficiency for experimental conditions matched their actual speed performance in all but one case: for the Fullscreen interface, participants in fact performed best on Jog. Squeeze and Click both induced inferior performance.

The participants' perception, however, was that they were less efficient with the Click modality than with Squeeze or Jog. According to our manipulation theory, one explanation is as follows: Jog provides the user with a steady stream of photos, because the jog dial is either in its resting position (no photo movement), or is pushed all the way up or down (single speed photo movement). In the Jog modality, photos thus either move or they don't. This pairing of simplicity with the convenience of single-handed operation may have led to the perception of high efficiency.

In contrast, Squeeze and Click each had one disadvantage on the participants' perception of efficiency. Squeeze requires the operator to consciously control the frame rate. While this flexibility in speed control is

presumably welcome in other cases, in our single-focus interface the control flexibility may have been perceived as a burden. Participants may have felt distracted by the need to manage the squeeze/frame rate control loop.

6.2 Performance Across Modalities

We had expected the low speed Click modality to work best with the high sense of place Thumbnail interface. However, in this experiment, Squeeze, Jog, and Click all had statistically similar search times in the Thumbnail interface. The reason for this is likely to be the same as that for Thumbnail's lowered physical strain. The time to assimilate a row of new photos during each screen update, as opposed to just one new addition, may have caused a leveling of the interaction speed across the modalities. Progress speed is dominated by the cognitive activity, rather than by the interaction physics. Conversely, when only one new image is exposed at a time, the progress speed is dominated by the input control rather than the cognitive activity.

We had originally expected Squeeze to do well with any interface because it allows the users to control transport speed continuously. Our results did not show this and in fact showed that the Jog modality performed well regardless of interface. In the Fullscreen interface Jog statistically outperformed the Squeeze and Click modalities. In the Parade interface, Jog performed the same as the Squeeze modality and statistically better than the Click modality. In the Thumbnail interface, Jog performed just as well as the Squeeze and Click modalities.

Utilizing our cognition and manipulation theories to elaborate upon the Jog results, we found that one possible reason for why Jog performed well regardless of interface was its adaptation to the advantages of each interface. For Fullscreen, if the user wanted to go fast and have the photos continue scrolling without any further effort on his part, Jog allowed the user to do this. For Thumbnail, if the user wanted to go slow and be sure that one scroll moved one row of photos, Jog allowed this.

It was more difficult for the other modalities to provide this range of control. With Squeeze, the user had a difficult time understanding exactly how much pressure moved how many photos. With Click, the user could only proceed as fast as he clicked, which required effort and was tedious. These modalities had a difficult time adapting to interfaces that they were not well suited to run with.

To conclude, we learned that Jog's compromise degree of control-effect coupling worked well. Its on/off clarity, combined with the advantage of continuous transport, helped search speed and reduced stress (over Click). Squeeze offered a broader range of control speeds, but its effect-control coupling seems to be too loose. It is always difficult to choose the value for an operating parameter to suit multiple users simultaneously. Jog's single transport speed is such a one-must-fit-all parameter. We had therefore expected the Squeeze modality to offer a clear advantage by allowing each person to control speed continuously. It seems, however, that the price we paid by loosening the control-effect coupling was too high.

We will examine in our follow-on work whether we can modify Squeeze to retain the necessary level of control-effect coupling and still provide high transport speed flexibility. One possibility will be to personalize the slope of the pressure/speed function automatically or by means of a short training run.

6.3 Perceived Lack of Reliability for Squeeze

The results show that most participants perceived Squeeze as the least 'reliable' of the three modalities. The squeeze sensor did not break down during the experiments, so the participants' understanding of 'reliability' was not technical in nature. Our theory is that the pressure sensor was overly sensitive. This sensitivity, while making the interface feel responsive, also led to frequent 'overshooting' past the target photo. Reversing display direction after such an error required the tapping of an on-screen button with the stylus. This action, which required the use of the second hand, may have been distracting enough that the overall experience of the overshoot error translated to a decrease in perceived reliability, rather than, for example, a penalty to the effectiveness score.

6.4 Physical Strain

We had anticipated that wrist strain would be a significant penalty for the Squeeze modality. Instead, participants found clicking to be the most strenuous interaction. One reason might simply be that clicking requires a larger number of physical actions for recalling a given number of photos in comparison to the jog dial or the squeeze sensor. Note, however, that for the Thumbnail interface all modalities were perceived to be statistically similar in strain. One explanation for this result is that each scroll notch of the photo display reveals a row of new photos in the Thumbnail interface, but only one new photo for the other interfaces. The Click modality may therefore have lost some of its strain-inducing effect because participants had to take breaks between each click to study the new group of photos.

7 Related Work

There are a number of commercial photo browser applications available for various handheld device platforms. We examined six of the most popular commercial photo browsers on a popular handheld software Website (www.handango.com). The browsers mainly use thumbnail views and Windows-Explorer-style folder views to browse through photos. Clicking or a jog dial can be used to control scroll bars. The browsers are geared towards browsing a small collection of pre-selected photos, rather than finding one specific photo in the user's complete image collection. Several projects have also studied different layouts for browsing photos on the desktop [1, 5, 6, 7].

Various modalities other than Click and Jog have been investigated to provide additional input and control to handheld devices, pressure sensor being one of them. Harrison et al. [8] detect contact with handheld devices using pressure sensors and demonstrate interaction techniques for scrolling, and for automatically detecting the user's handedness. Hinckley and colleagues [9] introduce and integrate a set of sensors, including a pressure sensor, into a handheld device. Their research provides several new functionalities engendered by the sensors, such as automatically powering up the device when the user picks it up, and scrolling the display using tilt. In the ComTouch project, Chang et al. [10] use a pressure sensor to translate hand pressure into vibration intensity between users in real-time. Tilt sensors have been explored for handheld devices as well. Rekimoto [11] uses tilting for menu selection and map browsing. Harrison et al. [8], Small & Ishii [12], and Bartlett [13] use tilt sensors to scroll through and select information on a handheld device.

Rapid Serial Visual Presentation (RSVP), electronically similar to the activity of riffling through the pages of a book to get a rough idea of the content, is based on the research result that humans have the ability to recognize the presence of a target image in as little as 100 milliseconds or less [14]. By presenting information sequentially and continuously, RSVP offers considerable potential when applied to electronic information spaces. Research into RSVP has a long history [15]. Various modes of RSVP have been investigated [16] and triggered strong interest from the areas of information visualization and cognitive psychology. Typical RSVP design modes include collage-mode, carousel-mode, floating mode, shelf-mode, and slide show (keyhole) mode. The advantages of RSVP on handhelds have also been studied [17-20].

8 Conclusion

We explored how three alternative interfaces interact with three different control modalities with respect to photo browsing on PDAs. We constructed a traditional thumbnail-based interface, an interface that placed one photo centrally onto the screen with a row of thumbnails above and below, and an interface that showed one full-screen photo at a time. We asked subjects to search for given pictures with each of these interfaces, using three different modalities: a squeezable input device, a jog dial, and a standard point-and-click facility.

We found that overall, the tri-state Jog modality, with its single speed, off/forward/reverse switch did very well. We had expected that the more flexible Squeeze modality, which allows users to control photo transport speed through finger pressure, would outperform the fixed-speed Jog. Our data indicates that this is not the case with the current implementation of the Squeeze modality. In terms of our design tradeoff

model we suspect that the control-effect coupling for Squeeze is currently not tight enough, even when a high sense of place user interface provides context within the collection. Our plan is to explore methods for retaining more control-effect coupling, while still providing good control flexibility.

Photo browsing on small devices poses many user interface design challenges. As the use of digital photography increases, the payoffs for addressing those challenges rise. Mobile photo management should be able to improve on traditional wallet pictures, for example by allowing more images to be portable. This portability is important not just for casual consumers, but also for a number of professionals who rely on image access in the field.

The potential for efficient and satisfying photo management on small devices is there, but additional science is needed to address open issues. These issues include not just linear scanning, but summarization, automatic labeling, effective search over photo collections, and controlled photo sharing. Augmenting these features with intuitive, highly efficient user interface design will empower users with increased portability and accessibility for their digital media collections. All of these are exciting areas to work in.

References

1. Bederson, B.: Photomesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, ACM Press (2001), 71-80
2. Paepcke, A., Wang, Q., Patel, S., Wang, M., Harada, S.: A Cost-Effective Three-in-One PDA Input Control. International Journal of Human-Computer Studies, Volume 60, Issues 5-6, (2004) 717-736
3. Furnas, G. W.: Generalized Fisheye Views. Human Factors in Computing Systems CHI '86 Conference Proceedings (1986), 16-23
4. Harada, S., Naaman, M., Song, Y., Wang, Q., Paepcke, A.: Lost in Memories: Interacting With Large Photo Collections on PDAs, Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries (2004)
5. Kang, H., and Shneiderman, B.: Visualization Methods for Personal Photo Collections: Browsing and Searching in the PhotoFinder. Proc. IEEE International Conference on Multimedia and Expo (2000)
6. Kuchinsky, A., Pering, C., Creech, M.L., Freeze, D., Serra, B., and Gwizdka, J.: otoFile: a Consumer Multimedia Organization and Retrieval System. FProc. CHI (1999), 496-503
7. Platt, J.C., Czerwinski, M., and Field, B.A.: PhotoTOC: Automatic Clustering for Browsing Personal Photographs. Microsoft Research Tech Report Number MSR-TR-2002-17 (2002)
8. Harrison, B., Fishkin, K., Gujar, A., Mochon, C., and Want, R.: Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces. CHI (1998), 17-24
9. Hinckley, K., Pierce, J., Sinclair, M., and Horvitz, E.: Sensing Techniques for Mobile Interaction. Proceeding of UIST 2000 (2000), 91-100
10. Chang, A., OModhrain, S., Jacob, R., and Ishii, H.: ComTouch: Design of a Vibrotactile Communication Device. DIS2002 (2002), 312-320
11. Rekimoto, J.: Tilting Operations for Small Screen Interfaces UIST96 (1996), 167-168
12. Small, D., and Ishii, H.: Design of Spatially Aware Graspable Displays CHI97 Companion (1997), 367-368.
13. Bartlett, J.F.: Rock n' Scroll is Here to Stay. IEEE Computer Graphics and Applications, 20(3), (2000), 40-50
14. Healey, C., Booth, K., and Enns, J.: High-Speed Visual Estimation Using Preattentive Processing. ACM Transactions on Computer-Human Interaction, June (1996)
15. Coltheart, V.: Fleeting Memories: Cognition of Brief Visual Stimuli, Cambridge, MA, MIT Press (1999)
16. Spence, R.: Rapid, Serial and Visual: A Presentation Technique with Potential, Information Visualization, Volume 1, Issue 1 (2002), 13-19
17. Lam, K., and Spence, R.: Image Browsing – A Space-Time Trade-off, Proceedings INTERACT97 (1997), 611-612
18. De Bruijn, O., and Spence, R.: Rapid Serial Visual Presentation: A Space-Time Trade-off in Information Presentation. Proceedings of AVI2000 (2000), 189-192
19. Derthick, M.: Interfaces for Palmtop Image Search, Proceedings of the Joint ACM/IEEE Conference on Digital Libraries, (2002), 340-341
20. Ware, C.: Information Visualization: Perception for Design (2000) Morgan Kaufmann