Evaluation of Image Browsing Interfaces for Smartphones and Tablets

Marco A. Hudelist  
Alpen-Adria-Universität Klagenfurt  
Institute of Information Technology  
9020 Klagenfurt, Austria  
marco@itec.aau.at

Klaus Schoeffmann  
Alpen-Adria-Universität Klagenfurt  
Institute of Information Technology  
9020 Klagenfurt, Austria  
ks@itec.aau.at

David Ahlström  
Alpen-Adria-Universität Klagenfurt  
Department of Informatics Systems  
9020 Klagenfurt, Austria  
david.ahlstroem@aau.at

Abstract—Smartphones and tablets are popular devices. As lightweight, compact devices with built-in high-quality cameras, they are ideal to carry around and to use for snapshot photography. As the number of photos accumulate on the device quickly finding a particular photo can be tedious using the default grid-based photo browser installed on the device. In this paper we investigate user performance in a photo browsing task on an iPad and an iPod Touch device. We present results from two user experiments comparing the standard grid interface to a pan-and-zoomable grid, a 3D-globe and a 3D-ring. In particular we are interested in how the interfaces perform with large photo collections (100 to 400 photos). The results show most promise for the pan-and-zoom grid and that the performance with the standard grid interface quickly deteriorates with large collections.

Keywords—Image Browsing, Mobile devices, Touchscreens

I. INTRODUCTION

Mobile devices like smartphones and tablets encourage users to take more photos and have them always available to show and share them with colleagues and friends. Many users have hundreds of snapshots on their devices [22]. At the same time a new challenge raises: how can image collections on mobile devices with touchscreens efficiently been browsed? The typical browsing interfaces on smartphones and tablets usually present thumbnail images of photos in a scrollable list that is sorted according to creation date. Using such grid-based lists with large image collections can become tiresome as typically much scrolling is involved to find a specific image. Usually, today’s devices also offer a facility to sort images according to geo-location in order to speed up search. However, if many captures are made at the same location within a short time span this approach becomes less useful. Sorting and clustering based on user-generated tags shows promise, but users typically do not apply any tags to their images themselves ([13], [5]).

Another option is to use content analysis [21] to abstract information and to use this information for search and browsing purposes. However, complex and rich analysis approaches are typically time-consuming and demand great processing power. Over the last years mobile devices saw an increased performance growth. Smartphones with two quad-core CPUs and multi-core GPUs are already a reality. Still, resources are limited on mobile devices compared to full-fledged desktop or laptop computers. Therefore, only rather simple and fast content analysis algorithms are feasible, which still can provide enough information to enhance mobile image browsing interfaces.

Smartphones and tablets additionally offer new interaction possibilities for users that are not possible on traditional computers. Multi-input low-latency touchscreens let users directly interact with interfaces and content. Sensors like build in accelerometers, gyroscopes, microphones, cameras or light sensors could be used to further improve the interaction, creating natural user interfaces. Although the resources of today’s mobile devices may not be sufficient to drive complex content analysis algorithms they have enough performance to enable visual appealing and rich browsing interfaces in 2D as well as 3D. Such interfaces can help users find wanted content more quickly and offer at the same time a pleasing experience.

We want to create a next generation of image browsing interfaces that use the capabilities of today’s mobile devices to its fullest and surpass the above described shortcomings of current mobile browsing interfaces. In our work we combine the advantages of mobile touchscreen devices with simple but efficient content analysis and 3D visualization techniques. We implement a set of early interface concepts for tablets as well as for smartphones and evaluate our designs in user studies. In the following, we first describe our approach for tablets. We present two 3D visualizations: a color sorted 3D image globe (Globe) and a color sorted 3D image ring (Ring). Both interfaces are evaluated in a user study and we show and discuss our results. In the next section we describe our work for small touchscreen devices like smartphones. We discuss changes we applied to the Globe and Ring designs on smaller displays. Further, we introduce a 2D pane-like interface (ImagePane) that operates at two zoom-levels (collection overview and zoomed view) and also uses color sorting. In a user study we compare all three and contrast them with a traditional grid-like interface like it can be found on most mobile devices. We discuss the results of the study and close with giving our conclusions and future work.
II. RELATED WORK

Improving the traditional scrollable list of images has been the focus of research as well as commercial companies. This is true for the desktop PC [9] as well as for mobile touchscreen-based devices like smartphones or tablets. Taking advantage of content analysis and different presentations on desktop PCs has also already been the focus of many works, especially in image and video retrieval [21]. In contrast to retrieval users typically do not formulate a search query in browsing scenarios. It is an interactive search and refining process. Users start with a coarse idea about what they are searching for (“I know it when I see it”). Over the length of the search process they refine it more and more until they reach what they are looking for. Another scenario is browsing for fun and exploration where users are not searching for a particular image. They just want to get an overview of the collection and enjoy the contents. Potential lies in the combination of rich 3D visualizations tuned with results from content analysis for mobile devices. For example, it was shown in [12] and [16] that color sorted thumbnail arrangements can improve image search. Other concepts range from exploiting a Fisheye-like view [2], a hexagonal arrangement [15], to 3D rings [18], 3D storyboards [19], cylinders and spheres ([11], [14]), which can be viewed and manipulated from the outside as well as from the inside. Other ideas use treemaps to organize images [10] or cluster very similar images with stacks [11]. Image search results are visualized on a smartphone using a 3D arrangement in [6]. The presentation should give the user a better idea about multi-dimensional result datasets. Regarding the design of the interfaces research exists that indicates that users prefer to be presented many images at once [13] and that they are comfortable with the usage of very small thumbnails [8].

III. TABLET INTERFACES

Ahlström et al. [1] compared different 3D globe interfaces for image browsing on tablets. They contrasted the different designs to a traditional grid-like interface in a user study. In the results they could show a statistical significant difference in search time for the Globe interface. In a next step we now compare their H-Globe interface with the Ring interface of Schoeffmann and Ahlström [17]. In both designs the placement of each individual image corresponds to the results of a pre-performed color based sorting. For this, the color sorting algorithm of Schoeffmann and Ahlström [16] was used. The algorithm sorts images based on their dominant hue level. The hue level is obtained through a 24-bin HSV histogram. Further, very bright images are placed at the beginning and very dark images are placed at the end of the sorted image list. The prototypes are implemented using Objective-C and the OpenGL ES 3D graphics library. As hardware an Apple iPad 3 WiFi with 16 GB of storage is used.

A. Globe

Each image is displayed on the surface of a 3D globe. The images are aligned in a grid-like pattern as shown in Figure 1. Images are taken from the color sorted list of images and applied to the cells on the Globe one by one, going north to south and west to east. Users can interact with the interface by applying touch gestures. Drag gestures rotate and tilt the Globe, similar to popular applications like Google Earth. Tilting is restricted to a maximum of 30 degrees up and down to avoid accidental and irritating turn-overs of the Globe. Pinch gestures can be used to zoom the view to get a better idea of the displayed images. Images can be selected for a bigger detail view by a single-tap. The Globe is then hidden and the selected image is displayed in larger size. To return to the Globe users can tap on a now displayed ‘Back’-button.

B. Ring

The Ring interface arranges the image collection in the appearance of a 3D ring. The base structure is a grid-like arrangement with 5 rows and 70 columns (see Figure 2). A user can manipulate the Ring by using dragging gestures (rotation) and pinching gestures (zooming). Further, a double-tap automatically zooms the current view to the back part of the Ring. Repeating this gesture returns the...
user to the initial view. The images of the pre-calculated color sorted image list are applied to the cells of the Ring from top to bottom going from west to east. Because of a slight tilting of the Ring in the initial state also images on the back of the Ring are visible. Therefore, more images can be displayed than what is possible on the Globe with similar image set sizes. The interface makes sure that the images at the back are not mirrored to reduce visual irritation.

C. Evaluation

For the evaluation, we performed a user study with 16 participants (15 male, one female). The average age was 27.1, five wore glasses. Fourteen stated that they would work 40 hours or more per week with a computer. Fifteen owned a smartphone with a touchscreen and four possessed a tablet.

Before the user test started all participants had to read through an instructional text that told them what the test was about. Further, they also had to read a short describing text before they started with an interface, which told them how to use it. Each participant had to find a minimum of 40 images per interface in a collection of 350 images. The collection was created by randomly drawing images from the Wang image set [23] and the IACC.1 TRECVID 2010 repository [20] as well as taking images from Flickr.

The target image was displayed at the beginning of each trial (see Figure 3a). When participants were ready they could start the trial by tapping on a displayed ‘Start’-button. This also started the time logging of the trial. To successfully complete a trial participants had to tap on the wanted image to display it in a bigger detail view. Next, they had to tap on a now visible ‘Accept’-button below the image (see Figure 3b) that also ended time logging for the trial.

![Figure 3. Start (a) and detail view (b) of a trial on the iPad.](image)

We implemented the detail view mode in order to be compliant to the default behavior of default grid interfaces like the photos application on iOS. Otherwise, the new interfaces would have had an unfair advantage and it would have biased the trial logging. Further, it gives participants a second chance to review their choice as they may mistakenly accept a similar looking but wrong image. When the accepted image was correct the whole screen displayed a green color for a few seconds. Otherwise, a red color was displayed. In case participants felt to give up on searching an image they could skip it by tapping on a ‘Skip’-button.

The image collection was structured in ten target groups with 35 images that are spatially near (e.g., similar to Figure 10). We use target groups in order to level out effects of different image difficulty experienced by different participants. From each group initially seven images as possible targets were chosen randomly (four default target images and three alternatives to cover skipping). A randomly pre-created target group list defined in which order target images were pulled from the target groups. The order in which target groups were prompted therefore remained the same for all participants for both interfaces. If participants chose to skip an image they immediately got the next target image of the corresponding target group. Should a target group be not able to provide another target image it was continued with the next target group in the target group list. Because of this, it could happen that different participants had to complete different amounts of trials, depending on how many trials they chose to skip.

After completing the test for one interface participants filled out a questionnaire and gave comments about their experience with the interface. In the questionnaires participants had to give Likert scale ratings based on the NASA TLX workload index [7] about ‘Mental demand’, ‘Physical demand’, ‘Experienced temporal pressure’, ‘Effort’, ‘Frustration’, ‘Fun’ as well as ‘Experienced support of interface visualization technique’, ‘Learning effect’ and ‘Experienced support of color sorting’. Participants had to choose a ranking between one and ten. A ranking of one was the best ranking for all questions except ‘Fun’, ‘Experienced support of interface’ and ‘Experienced support of color sorting’. When both interfaces had been tested and their questionnaires were filled out, participants were asked to give their final remarks and rank the two interfaces compared to each other. To level out learning effects the first interface was varied from participant to participant.

We report statistical significance of the factors interface, target group and interface × target group interaction as well as partial eta-squared ($\eta^2$). Eta-squared is a measure of effect size and should be interpreted as following: 0.01 represents a small effect size, 0.06 indicates medium and 0.14 indicates a large effect [3].

1) Results - Trial Distribution: For the Globe a total 691 trials were performed of which 618 trials were correct, 54 were skipped and 19 were wrongly selected as target image. For the Ring a total of 678 trials were performed of which 604 were correct, 39 were skipped and 35 were wrong. The total number of trials differ between the interfaces because users were allowed to skip trials and continue with alternative target images of the same target group.

2) Results - Trial Times: Analysis of trial times are based on correct trials only (618 for Globe and 604 for Ring). Search times were positively skewed. Therefore, we applied
a logarithmic transformation resulting in a distribution close to normal before analyzing the data. The geometric mean search time (i.e., the antilog of the mean of the log-transformed data) was 9.42 seconds for the Globe and 10.61 seconds for the Ring. A repeated measures ANOVA with independent factors interface and target group did not show a statistical significant main effect for interface \((F_{3,15} = 3.07, p = 0.1, \eta^2 = 0.17)\), but a significant main effect for target group \((F_{5,74.2} = 40.10, p < 0.0001, \eta^2 = 0.728\); Greenhouse-Geisser corrected) and a significant interface \(\times\) target group interaction \((F_{4.1,61.2} = 6.22, p < 0.0001, \eta^2 = 0.293\); Greenhouse-Geisser corrected).

In Figure 4 we see strong deviations between the interfaces in the Target groups 2 and 10. It is most likely that the significance of the interface \(\times\) target group interaction originates from these two cases. The difference can be explained in slightly different initial rotation angles of the Ring and the Globe. The Ring favored Target group 2 as it was more visible on the initial view than on the Globe. In accordance the Globe had an advantage in Target group 10, as the border between Target group 1 and 10 was centered. Therefore, participants had a better view of Target group 10 on the Globe. Otherwise, the differences of the interfaces were insignificant.

3) Results - Questionnaires: Both interfaces were rated rather equally in the questionnaires as can be seen in Figure 5 (note that lower is better for all measures except for ‘Fun’, ‘Support of Interface’ and ‘Support of Color Sort’). For each measure a T-Test was performed. In no case a significant difference between the interfaces could be measured. When asked for a favorite design after having used both interfaces, participants voted exactly 50:50 with eight voting for the Globe and eight voting for the Ring.

4) Results - Summary: In summary it can be said that both interfaces performed very similar. Under special circumstances the Globe seems to have performed a little bit better than the Ring (count of wrong trials, Target group 10) but those were partly caused by the setup of the user study. In normal use it is unlikely that users experience differences in search performance. This is also reflected in participants interface voting that was equally distributed among the two interfaces.

IV. SMARTPHONE INTERFACES

In our earlier user studies we focused on image browsing on tablets. In an follow-up study we want to expand our research to small scale touchscreen devices like smartphones. We adjust our previous Globe and Ring prototypes for the use on smartphone displays. Additionally, we add a new interface concept – the ImagePane (see Figure 6c). All three are contrasted in a user study with an implementation of a traditional grid-like image browser. For all interfaces the same color sorting algorithm is used as for the tablet interfaces. The prototypes are implemented using Objective-C as programming language and OpenGL ES as 3D graphics library. The used hardware was a fourth generation Apple iPod Touch with eight GB of storage that is comparable in terms of capabilities to an iPhone 4/4S.

A. Globe

The Globe implementation for smartphones is similar to the implementation for tablet screen sizes, which was described earlier (see Figure 6a). Count of rows/columns change accordingly to the set size to provide thumbnail sizes as large as possible. This is not only visually important but
to provide big enough hit targets on the smaller screen size. Otherwise it would have been too imprecise to select single images with a finger tap. Finally, the maximum rotation speeds are decreased as well as changing the maximum zooming levels.

B. Ring

Compared to the Ring of our earlier described study we fix certain aspects of the interface that were criticized by users. Screen usage was improved (see Figure 6b) by slightly changing height and position of the Ring. Further, it is no longer possible to see through the gaps between the images as this was noted as being very irritating. The interface was made less sensitive to accidental taps and new gestures are introduced. Users can perform swipe-up gestures to automatically zoom the view to the front of the Ring and swipe-down to return to the initial view. Double-tapping zooms the view automatically to the back of the Ring. The view of users is then positioned in the center of the Ring (see Figure 7). Double-taps directly on the Rings surface causes it to rotate in a way so that the touched region is centered in the zoomed view.

C. ImagePane

The initial view displays all images in the collection at once in a flat grid arrangement (see Figure 6c). The results of the color sorting algorithm are applied row by row from left to right. Thumbnail sizes change relative to the image set sizes so that all images can be shown in the initial view. When small image collections are presented the thumbnails are still big enough to make out details of the images. Users can directly select images for a detail view by single-tapping on the appropriate thumbnail. When large image collections are visualized the thumbnails of the images become quite small. In such cases it is very hard to make out details of the pictures but this is on purpose. Rather than directly choosing images from the initial view (which is still possible if users decide to) it has the purpose to give a good overview of the color sorting of the whole collection.

Users have the option to zoom into a certain region of the pane by double-tapping. In the zoomed view thumbnails become sufficient in size (comparable to a traditional grid-interface) to see enough details to choose the right image. In this mode it is also possible to scroll the pane horizontally as well as vertically to further explore the area. Another double-tap brings the user back to the initial overview of the collection.

D. Grid

The Grid resembles the traditional default image browser that is used on most available smartphones. It displays six square thumbnails per row. On one screen there is space for displaying three and a half rows of thumbnails (see Figure 6d), which is comparable to the default iOS image browser. Users can scroll through this list of thumbnails by applying drag gestures. It is only possible to scroll up- or downwards. Images are selected for a detail view by a single-tap.

E. Evaluation

To evaluate the four interfaces we performed a user study with 48 participants (23 male, 25 female). The average age was 24.2 years and 23 participants wore glasses or contact lenses. Self-reported computer use was on average 40.3 hours a week. All but four participants (44) owned a smartphone with a touchscreen and 16 owned a tablet. Participants were grouped in four groups of twelve. Participants were grouped in four groups of twelve. Each group tested one interface with four different image sets. Sets of 100, 200, 300, and 400 images were tested. New images were randomly drawn from Wang [23], TRECVID [20] and Flickr for each image set. Images of one set were
not reused in other sets but each set was used across all four interfaces. Testing a set of an interface consisted out of finding 60 target images. The target images were chosen randomly in the setup of the user study and remained the same for all interfaces for a given set. If participants needed longer than one minute for a trial the system automatically ended the trial (logged as ‘skipped’) and continued with the next trial. In contrast to our first described user study, no alternative images were automatically added and participants could not chose to skip a trial themselves.

To avoid fatigue effects, the study was split over two days. On each day, participants tested two of the four sets with their assigned interface. The order in which the different set sizes were tested was counterbalanced between participants and interfaces to avoid any possible order effects. On the first day participants were asked to carefully read an instructional text that explained the test procedure and how to use the assigned interface (presentation, use of touchscreen gestures, color sorting, etc.). After that they had to perform a training session. They then continued with the test trials. At the end of each day participants filled out a questionnaire in NASA TLX workload index style [7], which was structured exactly the same as in the first study.

A trial worked the same as in the first study: Participants were asked to memorize a displayed image. When they were ready they could start the trial by tapping on a ‘Start’-button and were transferred to the browsing interface. This also started the logging of trial time. To accept an image participants had to tap on its thumbnail to popup the detail view where the ‘Accept’-button was used to confirm the selection, which also ended the time logging of the trial (cf. Figure 3).

1) Results - Trial Distribution: In total, 11520 trials were performed. Each test candidate performed 240 trials by testing the assigned interface over all four sets. Each interface was therefore tested in 2880 trials. For the Globe interface 2505 trials were correct, 295 were automatically skipped and 80 were completed with accepting the wrong image. The Ring had 2452 correct trials, 287 trials were skipped and 141 trials were accepted with a wrong image. The ImagePane had 2449 correct trials, 279 skipped trials and 152 wrong trials. Finally, the Grid had 2696 correct trials, 142 skipped trials and 42 wrong trials.

2) Results - Trial Times: The analysis of trial times is based on correct trials only. Trial times were positively skewed and showed after a log-transformation a distribution near normal. A geometric mean search time aggregated over all set sizes showed 8.12 seconds for the Globe, 8.04 seconds for the Grid, 8.12 seconds for the ImagePane and 9.25 seconds for the Ring (see Figure 8b).

A mixed ANOVA with repeated measures and independent factors set size and target group (both inside subjects) and interface (between subjects) showed no significant main effect for interface ($F_{3,44} = 2.71, p = 0.056, \eta^2 = 0.156$) but a significant main effect for set size ($F_{3,132} = 398.94, p < 0.0001, \eta^2 = 0.901$), target group ($F_{5,7,252.6} = 179.81, p < 0.0001, \eta^2 = 0.803$; Greenhouse-Geisser corrected) as well as set size x interface interaction ($F_{9,132} = 4.51, p < 0.001, \eta^2 = 0.235$), target group x interface interaction ($F_{27,396} = 23.54, p < 0.0001, \eta^2 = 0.616$), set size x target group interaction ($F_{16.2,710.9} = 22.66, p < 0.0001, \eta^2 = 0.34$; Greenhouse-Geisser corrected) and set size x target group x interface interaction ($F_{81,1188} = 1.61, p = 0.001, \eta^2 = 0.099$).

Figure 8a shows that the Grid interface exhibits a rather linear increase of search time with increasingly large set sizes. In contrast, Globe, Ring and ImagePane show more of a leveling down curve with bigger set sizes. We speculate that if we would have tested additionally even bigger set sizes this trend would have continued.

Figure 9 shows geometric mean search times for the target groups across all interfaces and set sizes. A trend that is visible is that the Grid performs increasingly bad the higher the target group gets in contrast to the other interfaces. This can be explained with the increased scrolling needs for images at the bottom of the Grid. The other interfaces do not have such a problem and can provide faster search times for high target groups. A disadvantage of the Ring and the Globe interfaces at Target groups 3 and 9 is visible as can be seen in Figure 9.

This is likely caused by the fact that the images of these target groups resided at the sides of the 3D objects and were not or barely visible from the initial view (see Figure 10). A possible solution for this problem could be to avoid placing images at all at those positions. Especially the Ring interface
could be improved by splitting it into two halves, cutting away the hard to see area at the edges. Further, it can be seen that Ring and Globe perform very similarly, like it was the case in the first user study on tablets. Finally, also the rather equally distributed search times for all target groups of the ImagePane have to be noted. It seems that regardless of the position on the pane, the images could be found equally well by the participants.

3) Results - Questionnaires: Figure 11 shows the mean scores for each of the measures across all interfaces. For each measure a Kruskal-Wallis test was performed. No significant difference could be measured except for ‘Effort’ ($\chi^2(3) = 9.3, p = 0.026$) and ‘Frustration’ ($\chi^2(3) = 8.4, p = 0.039$). Pairwise comparisons performed using Dunn’s procedure [4] with a Bonferroni correction for multiple comparisons revealed significant differences for ‘Effort’ between Grid (Mdn = 5.0) and ImagePane (Mdn = 7.0) ($p = 0.029$) and for ‘Frustration’ between Grid (Mdn = 5.0) and Globe (Mdn = 7.0) ($p = 0.036$).

V. CONCLUSIONS AND FUTURE WORK

What can be said in general is that the Globe and Ring seem to perform very similar and have the same weak points. A conclusion that could be drawn from the results of the smartphone interface user test is that the ImagePane at least for small screen devices seems to be the most appropriate interface for color sorted image collections. It outperforms the Grid in all set sizes and the other interfaces in most set sizes. Further, the positive trend of the 3D interfaces with increasing set sizes has to be noted. Especially for bigger set sizes with 500 or more images it could be beneficial to create a combination of the ImagePane and the 3D interfaces.

Possible future work could involve creating a cover-flow-like visualization of multiple ImagePanes or a 3D ring of ImagePanes. Also the weak points of the Ring and the Globe interface at the sides should be considered. An improved Ring that is split into two arcs with less curvature could show an improvement in search performance. With such a design, the problem of not or poor visible images at the left and right side areas of the Ring would be eliminated.

We also noticed in our evaluations that it is possible that two images of the same target group have strong differences in their geometric mean search times. We saw this for specific images across different users. Thus, it could be beneficial to explore ways to rate or weight images according to their “search difficulty” and to consider such ratings in future evaluations.
ACKNOWLEDGMENTS

This work was funded by the Federal Ministry for Transport, Innovation and Technology (bmviit) and the Austrian Science Fund (FWF): TRP 273-N15 and the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF), supported by Lakeside Labs GmbH, Klagenfurt, Austria.

REFERENCES


