Brush-and-Drag: A Multi-touch Interface For Photo Triaging

Seon Joo Kim,[1] **Hongwei Ng, Stefan Winkler**
Advanced Digital Sciences Center (ADSC)
University of Illinois at Urbana-Champaign
Singapore 138632
stefan.winkler@adsc.com.sg

Peng Song, Chi-Wing Fu
School of Computer Engineering
Nanyang Technological University (NTU)
Singapore 639798
cwfu@ntu.edu.sg

ABSTRACT
Due to the convenience of taking pictures with various digital cameras and mobile devices, people often end up with multiple shots of the same scene with only slight variations. To enhance photo triaging, which is a very common photowork activity, we propose an *effective and easy-to-use brush-and-drag interface* that allows the user to interactively explore and compare photos within a broader scene context. First, we *brush* to mark an area of interest on a photo with our finger(s); our tailored segmentation engine automatically determines corresponding image elements among the photos. Then, we can *drag* the segmented elements from different photos across the screen to explore them simultaneously, and further perform simple finger gestures to interactively rank photos, select favorites for sharing, or to remove unwanted ones. This novel interaction method was implemented on a consumer-level tablet computer and demonstrated to offer effective interactions in a user study.

Author Keywords
Digital photo collections; User interaction

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User interfaces; I.4.6 Image Processing and Computer Vision: Segmentation

INTRODUCTION
Digital cameras have become affordable and nearly ubiquitous with their integration into mobile devices, enabling people to take pictures and videos almost anywhere and anytime. Due to this convenience and the low cost of digital memory, people often take multiple shots of the same scene, and then select the best picture(s) from the set afterwards. This is especially common for photos that involve people, e.g., family photos with energetic babies and kids, or photos of certain important events such as weddings or graduation ceremonies. In these cases, we want to capture the best moment with the most natural and touching faces with the right gazes for all the subjects, and share our favorite photos with family and friends. Many modern consumer-level cameras (even basic models) now offer features that facilitate this behavior, such as powerful continuous burst-shooting modes. As a result, users often end up with a large number of photos, which later require tedious photowork (photowork refers to activities people perform with their digital photos after capture but prior to end use such as sharing [16], such as assessment, selection, edit, organization, and annotation). Among these photowork activities, one of the most common and time consuming tasks is *photo triaging* or photo sorting.

Existing photo software provides very limited computational or interface support for photo triaging. For example, when using Adobe Photoshop, Google Picasa, or Apple iPhoto, this basic task still relies on a primitive interaction method: flipping through the photos and viewing them one by one. The user has to go back and forth among the candidate photos to look for the relevant image features, to compare their differences, to select his/her favorite(s), and/or to remove the undesirable photos. Given the difficulty of recalling differences between images, this simple method becomes ineffective and progressively more time consuming when there are more photos in the set. On top of this, one typically needs to go back and forth a number of times to check or verify the differences. Using thumbnail images as an alternative does not work well either because the relevant image features to be compared are often too small in a thumbnail view, thus relevant details such as facial expressions are not easily recognizable. This is especially problematic on mobile devices such as laptops or tablets with limited screen space and resolution.

This paper introduces a novel and effective user interface to improve the process of photo triaging: browsing, sorting, and selecting photos from a candidate set that contains multiple images of the same scene with similar content. We are particularly interested in developing novel multi-touch-based interactions for photo triaging on a mobile tablet computer such as the Apple iPad 2. This family of platforms has become highly popular among the general public as evidenced by the iPad's record-setting adoption rate.[2] Hence, devising interactions for mobile devices is becoming increasingly important [6,13,17,19,22], yet photo browsing interfaces for mobile devices are still based on rather primitive interaction methods.

MobileHCI'12, September 21-24, 2012, San Francisco, CA, USA.

[1] SJ Kim is now with SUNY Korea.
[2] http://www.appleinsider.com/articles/10/10/05/apples_ipad_proclaimed_to_have_fastest_adoption_rate_ever.html

Specifically, we propose a novel *brush-and-drag* interface, which works as follows. After grouping a set of similar photos interactively in a thumbnail view, the user can *brush* to mark an area of interest on one of the photos in the set directly with his/her finger(s). Then, our tailored segmentation engine automatically determines the corresponding image elements across all the photos in the group. After that, the user can *drag* with one or two fingers to bring out the corresponding image elements and spread them out on the tablet surface, thus revealing the related area of interest on all photos simultaneously. Finally, the user can perform various simple finger gestures to interactively rank and re-order the photos, select the favorite for sharing, or remove some of them.

This is a focus+context design: our interface allows the user to selectively choose any area/object of interest by brushing (focus) while retaining the overview photo (context). Thus, the photo triaging process becomes more flexible and user-centric. We implemented our interface design on the Apple iPad 2 and evaluated it with a number of users. Results show that our interface outperforms the conventional photo triaging method (browsing photos by flipping) in both objective and subjective measurements. As demonstrated by our user study, participants were able to select favorite photographs from groups of similar images more quickly with our interface.

RELATED WORK

Photo browsing and triaging have been active research topics in human-computer interaction. Several user studies have been performed to analyze how people manage their digital photo collections. Rodden and Wood [24] suggested that digital photos are easier to manage than their non-digital counterparts. They found that simple browsing features were much more important to users than more advanced features like content-based image retrieval or voice annotation by speech recognition. Later, Kirk et al. [16] introduced the notion of "photowork" and similarly concluded that image searching features have much less relevance than new ways of browsing photos.

Apted et al. [3] introduced a multi-user, multi-touch, collaborative interface for sharing photos on a tabletop display and studied various design issues when using it with young adults and older users. Additionally, social aspects of sharing photos online have been studied by Miller and Edwards [20] and Ames et al. [2], who investigated various requirements for mobile photoware, including image capture, upload, annotation, archiving, sharing, and viewing. These user studies again suggested that basic features for photo browsing such as triaging, displaying thumbnails, and sorting are much more important to users than advanced features.

A large amount of research effort has been devoted to improving the user interaction and visualization capabilities for digital photo browsing. Kang and Shneiderman [15] introduced a scatter plot thumbnail display and a drag-and-drop interface as a tool for searching and browsing photos, while Bederson [5] proposed a zoomable image browser using quantum treemaps and bubblemaps. Platt et al. [23] developed an overview+detail design, where the overview of the photo collection is generated by an image clustering algorithm, and the details are shown as an ordered list of photos. Graham et al. [12] applied a cluster analysis of the times when photographs were taken in order to summarize the photo collection. Huynh et al. [14] proposed a zoomable layout for displaying large photo collections by trading off screen space utilization to more effectively convey the temporal order of pictures. Shneiderman et al. [27] presented an integrated interface for annotating, browsing, and sharing photo collections, supporting exploratory search needs of personal digital photo users. Other photowork activities studied in the HCI research community include photo annotation [9, 28], photo sharing [11, 20], and photo storytelling [10, 31].

While thumbnail view or grid layout approaches allow us to view and sort multiple photos simultaneously, we argue that detailed features such as faces could be too small to see clearly, especially on tablets or handheld mobile devices, whose display sizes are relatively smaller. In addition, users do not have the ability to flexibly select and examine the details. In comparison, our interaction design provides the users with an interface to mark up feature details and compare them more flexibly and effectively with the support of multi-touch gestures and an interactive image segmentation engine.

Interfaces for browsing and exploring photos have also been studied in the computer graphics community. Snavely et al. [29] introduced an interactive 3D photo browsing interface, which puts photographs in a virtual 3D space relative to their corresponding camera viewpoints. Ballan et al. [4] extended this 3D interface design for exploring videos. Agarwala et al. [1] proposed the idea of digital photomontage, where image regions from a set of photos can be interactively composited seamlessly into a desired target image. Such a technique was later integrated into various commercial products, including Adobe Photoshop's Groupshot and Microsoft's PhotoFuse. The motivation of this method is similar to ours, i.e., helping the user to come up with a good representative photo from a set. However, their work focused on merging and compositing image elements from different photos while users still need to look for good image features among a photo set before selecting the preferred image parts for photo composition. Our proposed system on the other hand provides an interactive solution for efficiently comparing user-desired image features among photos, and it can complement the photomontage framework particularly in the triaging and selection process.

DESIGN ISSUES

We propose the brush-and-drag interface as an effective means for photo triaging (Figure 1). The interface is targeted at tablet/handheld computers and general-public users. Before presenting it, we enumerate a number of design issues incurred in this setting.

- **Easy-to-use, familiar interaction gestures.** We aim to develop easy-to-use and intuitive interaction gestures that are familiar to general-public users and consistent with the way touch interfaces have been developed in common surface computing applications.

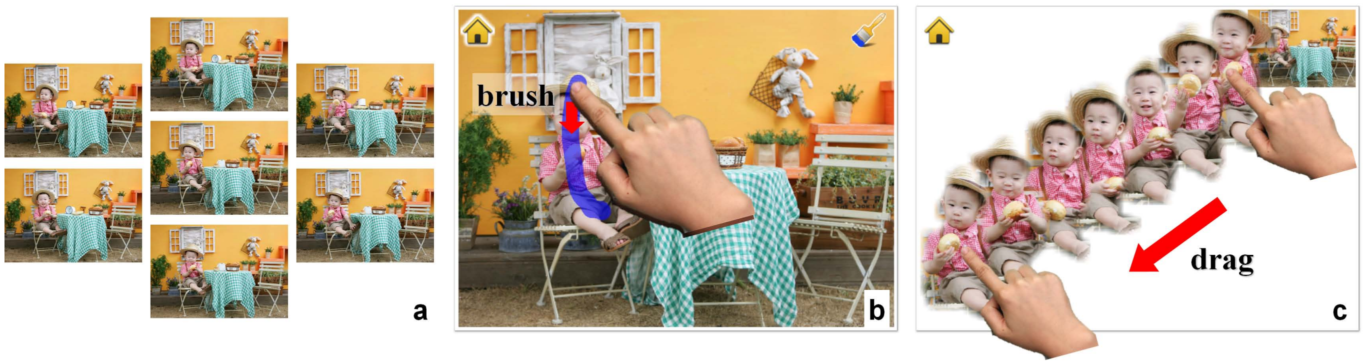

Figure 1. The brush-and-drag interface. (a) A given group of similar photos. (b) We use a natural *brush* gesture with our finger(s) to mark an area/object of interest in one of these photos. Our tailored segmentation engine quickly segments the corresponding image elements in the photo group. (c) We can then *drag* the segmented image elements to spread them out for better examination. This user-centric design allows the user to decide the size and the area of interest. It also maintains the overall photo context while enabling the user to focus on the selected area of interest for comparison and photo triaging (focus+context). Note that the normally black background of the tablet display was changed to white for printing purposes.

- **User input imprecision.** Since we rely on finger contacts on the multi-touch surface to provide input to the system, our system should not assume precise touch input due to the fat finger footprint, similar to other touch interfaces.
- **Visual feedback.** Since photo triaging requires a series of interaction stages, our interface should strive to inform the user about the changes in the interaction stages and actions they perform via appropriate visual feedback.
- **Focus+context interaction.** Due to the limited screen size and resolution available on tablet computers, we have to consider both the photo overview and the focus on details in the user interface. Hence, we take a focus+context approach to the interaction design, which allows users to stay within the overall photo context while focusing on and interacting with specific areas of interest.
- **Efficient computation.** Since the interface is intended to run interactively on a tablet computer, we should minimize the memory and computational overhead required by any supporting computation and algorithms in the system, in particular the image segmentation engine.

INTERACTION SCENARIO

The interaction scenario of our photo browsing tool consists of the following four stages:

- **Stage 0:** Before entering the main brush-and-drag interface, which starts from stage 1 below, the user first groups a set of similar looking photos in the *thumbnail view* of a photo album, see Figure 2.
- **Stage 1:** Being presented with the first photo from the selected group, the user can then *brush* to mark an area/object of interest on it directly with his/her finger(s), see Figure 1(b). This triggers our image segmentation engine to automatically select the corresponding image elements from the other photos in the set.
- **Stage 2:** The user can *drag* out the corresponding image elements with his/her finger(s) to arrange them along the dragging path for quick and intuitive comparisons in a focus+context style, see Figure 1(c).
- **Stage 3:** The user can further perform *a variety of easy-to-use photo triaging activities*, by manipulating the segmented image elements with his/her finger(s) to rank and re-order the photos, to select the favorite(s) for sharing, or to delete some of them.

We now present the details of each of these stages below.

Stage 0: Grouping Similar Photos

Like conventional digital photo browsers, our system starts with a thumbnail view of the photo collection. While comparing features and object details with thumbnail images can be difficult, the user can still identify similar photographs in this view. Thumbnails are arranged according to the time when the photos were taken by checking the photos' meta information. Hence, similar photos are automatically placed close to one another. Additionally, we introduce gaps between neighboring photos when the difference between their timestamps is relatively larger than others as a simple visual guide for the user; of course, other forms of visual indicators can also be envisaged.

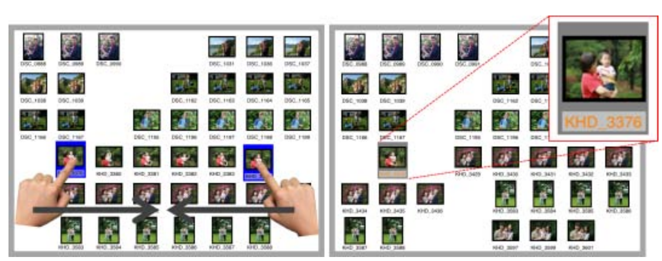

Figure 2. Stage 0: Thumbnail images are automatically arranged according to their photo-taking timestamps. (Left) The user can group photos by pressing on the first and the last photo of the set with two fingers, and then collapsing them into a batch. (Right) The thumbnail for the resulting photo batch is presented in a different style.

Our system provides an easy way of grouping photos into a batch as follows:

- First, the user places one finger on the first photo and another finger on the last photo in a group, see Figure 2 (left).

- Then, the user can do a pinching gesture to collapse all photos in-between into a batch.
- The icon for the newly-created photo batch is then presented in a different style, see also Figure 2 (right).
- Finally, a simple one-finger tap on a photo batch takes the user to the next interaction stage, *brushing*.

If the user taps on a single photo thumbnail instead of a batch icon, a full view of the photo is presented, and the user can browse the collection using the standard flipping/scrolling gesture. In this way, our interface works just like a conventional photo browser.

Stage 1: Brushing

Upon tapping on a photo batch icon in the thumbnail view, we enter the brushing interaction stage, which starts by showing a full view of the first photo in the batch. There are two modes of interaction in this stage, namely *scrolling mode* and *brushing mode*. When the interface enters this stage, it starts in scrolling mode, which allows the user to browse through different photos using the standard flipping gesture. Brushing is not allowed in this mode because it may be confused with this gesture. For disambiguation, we provide a button on the top-right corner of the interface in the form of a paint brush icon, which allows the user to switch between scrolling and brushing modes. There is also a home icon shown on the top left, which returns the user to the initial thumbnail view at any time, cf. Figure 1(b).

After the user taps the brush icon, it changes to a blue paintbrush as a visual metaphor to indicate that the brush is now ready to use. Now, the system enters the brushing mode. The working scenario is as follows:

- First, the user marks an area/object of interest by freely brushing with his/her finger(s), which is one of the core features of our brush-and-drag design. Multiple fingers can be used simultaneously for brushing by taking advantage of the increased degrees of freedom provided by the multi-touch surface. This feature allows the user to intuitively deal with varying object sizes without having to control the diameter of the brush stroke explicitly. Note that an area of interest does not have to comprise just one object, as the user may be interested in focusing on multiple objects that are located next to each other, as shown in Figure 3.
- Once the finger(s) is/are lifted from the multi-touch surface, our system initiates the segmentation engine to automatically determine the corresponding image elements across all photos in the batch. This segmentation engine is another key contribution of this work. Users only need to mark the area of interest in one of the photos. The engine can then determine the corresponding image elements across all photos in the batch. This user-driven segmentation further enhances the interaction capabilities and personalization of photo browsing. Details of our interactive segmentation engine are described in the Appendix.
- Once the segmentation is complete, the system automatically switches to the *pre-dragging mode* of stage 2 (dragging stage), and the top-right brush icon disappears, since we are no longer in the brushing mode. At the same time, the segmented element will be highlighted against a dimmed background, see Figure 3. If the user is not satisfied with the segmentation result, he/she can go back to brushing mode by double tapping the background. This removes the highlighting and makes the brush icon re-appear on the photo. The user can then start over with brushing. Alternatively, the user may tap the brush icon to revert back further to the scrolling mode of this stage.

Figure 3. Interactive segmentation by brushing: (top) brush strokes and (bottom) segmentation results highlighted by dimming the background. Brushing with multiple fingers naturally allows the user to widen the area of interest, which facilitates marking larger objects.

Stage 2: Dragging

The dragging stage has two modes: the *pre-dragging mode* followed by the *dragging mode*. We enter the pre-dragging mode after the segmentation. In this mode, the segmented image elements are highlighted against a dimmed background, as shown in Figure 3.

The user can now perform a dragging gesture with one or two fingers, which is a natural and intuitive action to spread out the segmented image elements along the dragging path. Two possible dragging gestures are supported in the pre-dragging mode, as can be seen in Figure 4:

- One-finger drag, in which the system automatically estimates the scale factor for the segmented elements, aiming at distributing them efficiently along the dragging path.
- Two-finger drag, where the size of the image elements is scaled relative to the distance between the two fingers, and image elements closer to the fingers are scaled relatively larger. This view can be helpful when sorting photos later on, because the preferred photos placed near the end of the dragging path are larger.

Using the dragging action described above, we can spread out the segmented image elements and present them all together over the given photo, following the idea of focus+context. However, the image elements presented at this point may be too cluttered, thus affecting the focus+context visualization as well as the subsequent interactions. We propose two mechanisms to resolve this issue:

- Once the finger(s) reach the tablet boundary (within a margin) as in Figure 5 (step 2), our interface automatically and progressively reduces the size of the context photo toward

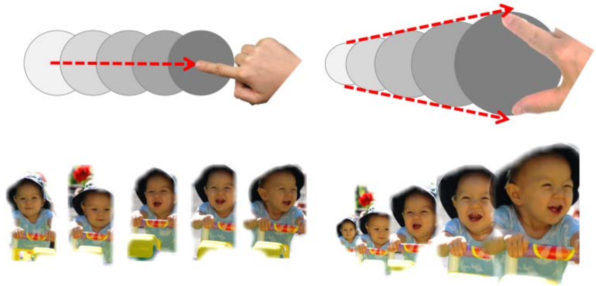

Figure 4. Dragging gestures. (Left) The one-finger drag uses the same size for all image elements. (Right) The two-finger drag scales the image elements similar to perspective viewing.

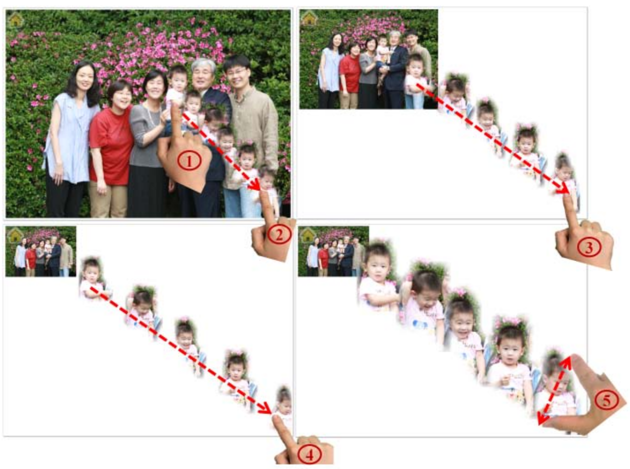

Figure 5. Automatic resizing allows us to interactively balance context and detail. After the dragging finger(s) reach the tablet boundary, the size of the context photo reduces progressively (see steps 1-4), making space for the image elements. The user can also perform a two-finger pinch to further adjust the size of the image elements (step 5).

the opposite direction (Figure 5, steps 3 and 4), making additional room for the image elements. Auto-resizing stops when the finger(s) are lifted up from the tablet surface. This mechanism allows the user to interactively balance the utilization of screen space between the main context photo and the details (segmented elements).

- Occasionally, there may be a large number of similar photos in a batch (perhaps 10 or more). In this case, the displayed elements would be either too small for effective comparison or too cluttered if the user tries to resize them (see Figure 6 (left)). To resolve this issue, we introduce the idea of a general dragging path, which allows the user to drag and spread the image elements along an arbitrary path rather than a straight line (the default) to better utilize screen space. Figure 6 illustrates this mechanism. The general path drag can be initiated after the default straight-line drag by simply pressing the context photo area with two fingers.

After the initial drag, the user may perform:

- A two-finger pinch to additionally adjust and fine-tune the size of image elements.

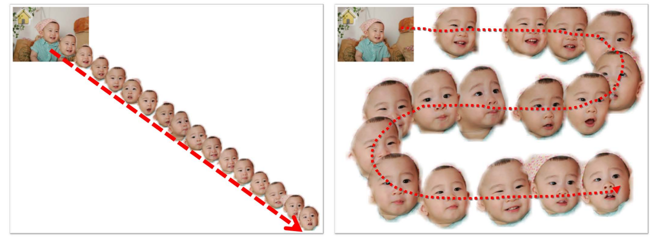

Figure 6. General path dragging. (Left) With a large number of images (16 in this example), a straight line arrangement of elements may not be effective. (Right) Users can drag the elements along a general path, which utilizes screen space more efficiently and allows more elements to be displayed at a larger size.

- A tap on an image element to switch the current main context photo to the photo associated with the tapped image element, enabling photo browsing by detail selection.
- A long press of 0.3 seconds on any image element, which takes the user to the triaging stage. This is different from a re-dragging gesture, which consists of a one-finger drag starting from the first element on the path to re-adjust and move an existing dragging path.
- A double tap on the main context photo takes the user back to pre-dragging mode, where the segmentation result is shown on the full-size view.

Stage 3: Photo Triaging Activities

In this stage, our user interface presents the overall context, which is the main context photo at the corner, as well as the focus/detail, which is an array of dragged image elements corresponding to the area/object of interest the user intended to explore earlier. This focus+context style of presenting the image segments provides not only the views on the user-marked image elements, but also interactions on these elements, making the photo triaging process more efficient by directly working with the focused elements.

Once we spread out the image segments along a dragging path, a long press on an image element enables the user to select that element and enter the *photo triaging stage*. The selected image element becomes more transparent to visually indicate the selection. The user can then perform the following photo triaging activities, as illustrated in Figure 7:

- Drag the selected image element to the star icon located near the top right corner of the interface to mark the associated photo as a favorite. A star will also appear in the top-right corner of the selected image element to indicate success.
- Drag the image element to one of the social network icons on the lower left to share the associated photo with friends.
- Delete a photo by dragging the associated image element to the recycle bin icon on the lower right. After this action, the related image element will disappear, and the associated photo will be moved into the recycle bin.
- Rank or sort photos interactively by dragging the selected image element and dropping it to a preferred location along

Figure 7. Photo triaging activities. (a) An image element is selected by a long press, and the selection is indicated by making it transparent. (b) The user can drag the selected element to one of the two social network icons to share and post it, to the star icon to mark it as a favorite, or to the recycle bin icon to throw it away. (c) The user can also rank or sort the images by interactively moving them before or after other image elements in the sequence. (d) The resulting photo order after sorting.

the path. The positions of other image elements will be rearranged accordingly.

- Go back to pre-dragging mode by double-tapping on the main context photo.

USER STUDY

To evaluate the effectiveness of our proposed interface, we conducted a user study, in which we compare our interface with the standard photo browser provided by Apple on the iPad 2. The reason behind this choice is that most existing photo applications on the iPad focus on enhancing and re-touching photos, whereas the basic photo browsing function remains very similar among different photo applications, without specific features for photo selection and sorting. The standard browser offers a thumbnail view as well as a simple browsing function, where the user flips from one photo to the next by a finger drag gesture. Note that our interface is also designed and developed to run on Apple's iPad 2.

Our evaluation is task-based. In each task, the participants were asked to select a favorite photo from a group of similar pictures, and the time taken by each of them to do the selection was recorded, so that we can use these measurements to evaluate the effectiveness of the two interfaces.

Test Photos

We compiled four sets of photos, which contain 4, 6, 8, and 10 similar images, respectively. We chose photo collections of a baby in various settings as the test sets, because this represents a common type of family photo, where one would usually take a large number of similar photos. Differences between photos in a set include facial expressions, the subject's poses, and modest changes in the background due to the camera motion. We have a progressively increasing number of photos in these sets, hoping to see how the number of photos may affect the task performance.

Participants

We recruited 24 participants for the user study and randomly divided them into two equally-sized groups (G1 and G2). G1 performed favorite photo selection on the four sets using Apple's standard photo browser followed by our interface, whereas G2 used the two interfaces in reverse order. There were no significant differences between the groups in terms of age and level of familiarity on using tablet computers and multi-touch interfaces. G1 comprises 8 males and 4 females, from 24 to 47 (mean 30.8) years old, and G2 comprises 7 males and 5 females, from 25 to 45 (mean 31.8) years old.

Notice that the same pre-defined sets of test photos were used for both G1 and G2 in order to minimize potential bias in the contents. When presented to the participants, the photos within each set were randomly arranged for each user and each interface.

Procedure

First, participants were asked to fill out a form with basic information including gender, age, and familiarity of using tablet computers. Then we briefly explained the procedure of the study, instructed them how to use the two interfaces, and let them practice with both (about a minute for Apple's browser and about 5 minutes for the brush-and-drag interface) so they could familiarize themselves with the two interfaces before doing the tasks.

After that, the participants started to perform the task of selecting their favorite photo from each test set. The participants in G1 were asked to select photos using the Apple's browser first. They started with the 4-photo set and ended with the 10-photo set. The participants in G2 were asked to perform the same task using our brush-and-drag interface first with the same order of sets. During the task, we measured the time taken by each participant to select a photo from each set. After the experiments, participants were asked to fill out a questionnaire and rate various aspects of the two interfaces.

Results

For the time measurements, we only use the results from the first round of selection before the participants switch to the second interface, which is our interface for G1 and Apple's standard photo browser for G2. The results from the user study confirmed that the participants indeed completed the selection task faster in the second round. Some reported to us that they could remember some of the photos they saw in the previous round, and thus could simply search for the same favorite photo they selected previously. However, we also encountered one participant who selected different favorite photos with different interfaces, reporting to us that his mood during the selection affected his choice. The reason why we require all participants to try both interfaces is that we also ask participants to rate both interfaces.

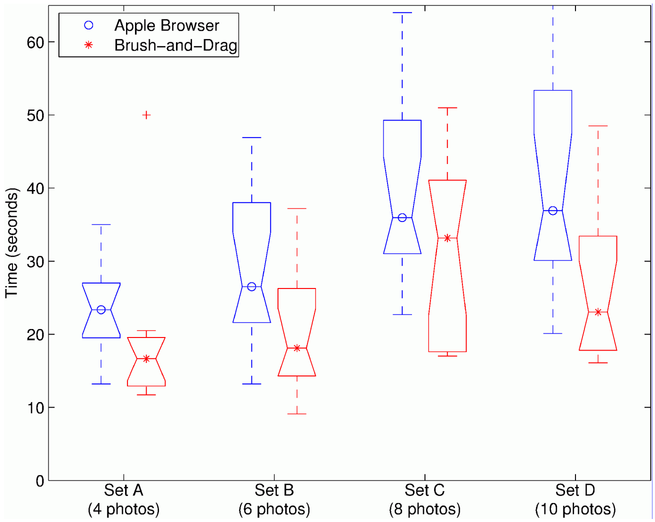

Figure 8. Box-and-whisker plots[3] of the time taken for favorite photo selection for each test set and group. Participants were able to select their favorites more quickly using our interface (red/stars) compared to Apple's (blue/circles).

Figure 8 shows box-and-whisker plots of the time participants took to select their favorite photo. For all test sets, photos were selected more quickly using our interface. Our interface can save the participants about 37% of selection time for the photo set with 10 images, and about 30% on average for all sets. Unfortunately we did not have enough test sets to investigate effects of image content.

Ratings

For the subjective ratings such as usability, we use ratings of both interfaces from both G1 and G2. After finishing the given tasks, participants were asked to rate the two interfaces on various aspects on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The results are presented in Figure 9.

We expected our brush-and-drag interface to facilitate the photo comparison, which is reflected in the ratings from the users. Somewhat surprisingly perhaps, participants found the brush-and-drag interface almost as easy to learn and use as the Apple photo browser, although a few did mention that ours was harder to pick up. In addition, they viewed our interface more fun to use. They liked the interactive selection of the area of interest as well as the handling of a group of image elements via dragging. Most importantly perhaps, we outperformed the conventional browser in terms of ease of photo triaging (selection and sorting), which is the main goal of this paper. All rating differences are significant at a level of at least 0.001 according to a paired t-test, except for ease-of-use ($p = 0.0128$).

[3] The boxes indicate the 25th and 75th percentile of the data, and the vertical line inside the box represents the median. The dents on the boxes indicate the 95% significance levels. The whiskers show the data within 1.5 times the interquartile range of the lower and upper quartiles, respectively; data beyond this range is plotted individually as crosses.

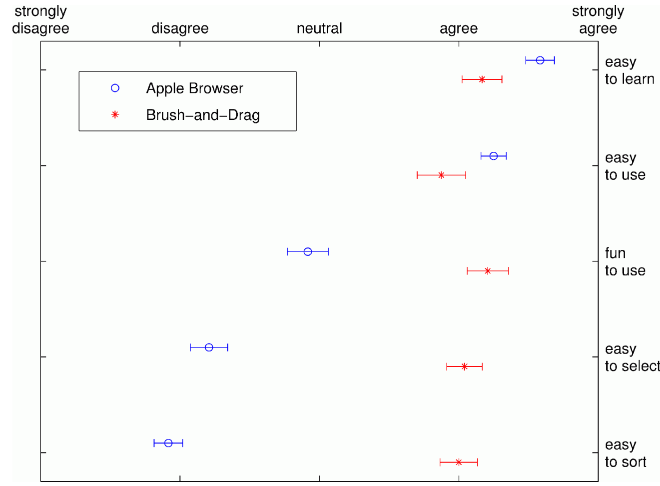

Figure 9. User ratings (means and 95% confidence intervals) of various aspects on a scale from 1 (strongly disagree) to 5 (strongly agree). Our system (red/stars) fairs well in all categories.

DISCUSSION

The questionnaire asked participants about their approach to photo-taking and selection. The large majority said they regularly took 3 or more photos of the same scene in order to capture a good shot, with some taking up to 8 or 10 in certain situations. Multiple shots were considered most useful for people, especially children and group shots, as well as moving objects. People's expressions were almost unanimously cited as the main criterion for choosing photos, with image clarity (i.e., sharpness, color balance, lighting, etc.) second, which matches findings from previous studies on image appeal [26].

Most participants performed the given tasks as we had intended. However, a few participants also came up with interesting alternative approaches. For example, in the segmentation stage, rather than brushing directly on the area of interest as they were shown in the training, some participants attempted to circle or outline it instead. Our segmentation engine is not designed for that, and the resulting image elements were typically much larger than the participant expected. However, it can be considered as an equally intuitive way of selecting an area/object of interest, and the segmentation engine could be adapted to handle both types of input, possibly by recognizing high-level stroke semantics and image features. Other alternatives to brushing can also be envisaged, e.g., selection by rubber-banding, or simply clicking on different image features. These may be more suitable for traditional PC platforms with mouse input rather than touch screens.

Another example is the use of the segmented image elements for the favorite photo selection task. Again, most participants followed the training and relied primarily on the different elements to make their choices: they enlarged the image elements and reduced the size of the context photo. Some participants however decided to use small image elements while keeping the context photo at the original size. To them the image elements served as a form of thumbnails, with which they could quickly switch between the different full-size images. Compared to the conventional photo browser, where flipping can only be done between neighboring images, the random-access capability we provided still enabled the participants to make their selections more quickly.

The placement of the segmented elements by dragging them along a user-specified path could also be facilitated by the system automatically choosing the optimal distribution of elements on the screen (depending on their content and shape) to maximize screen space usage. The user could subsequently make adjustments to the placement if necessary.

Most people brushed either the face(s) or person(s) in the images as the area of interest. This can be expected based on the comments from participants we received as well as other photowork studies, which show that people or faces and their expressions are among the most important features and selection criteria. It also led some participants to suggest that the system could offer automatic pre-segmentation of faces or people in the photo for comparison; the user could then choose to accept or override the suggestion.

Other comments revolved around certain design choices, such as the placement of buttons on the interface, or the need to perform a long press to initiate photo triaging functionalities, which could still be optimized. For some functions, soft buttons may be more intuitive than double-clicks. Phase 0 could be automated using photo timestamps and some image similarity criteria.

Some other limitations of the system still need to be addressed as well. The segmentation algorithm did not always give participants the desired result. It also requires the object/subject of interest to be in a similar position and location across photos, even though some variations are possible. More advanced tracking or identification methods would have to be employed to detect substantial movements. Furthermore, the user can only select a single contiguous area of interest; while this may be sufficient for the majority of cases, it would be interesting to extend the interface to facilitate the comparison of multiple disjoint areas in a photo.

CONCLUSIONS

This paper introduced a new interaction design, a *brush-and-drag* interface, for supporting efficient photo triaging. Users can interactively select objects of interest to facilitate common photowork activities on a conventional tablet computer. By a natural brush gesture, users can flexibly mark any area of interest in a photo, and by further dragging out the marked object, they can browse and explore corresponding image elements across a group of similar photos at the same time.

Compared to the traditional way of flipping back and forth between photos, which is a rather tedious and time-consuming process that requires our eyes to constantly search and compare, our interface features a user-centered design, offering the flexibility to choose where to brush and what areas to mark. Furthermore, after corresponding elements are dragged out, we offer a novel user experience for photo triaging activities, allowing users to directly make use of the segmented elements for the selection of favorites as well as sharing and deleting photos. Finally, our focus+context design allows users to maintain an overview of the entire photo context while exploring the details of interest.

Future Extensions

We mostly focused on family photos such as portraits, because this is one of the most common situations where people take multiple shots (as was also confirmed by the participants in our study). It will be interesting to develop a similar triaging system for landscape photos, which bring different challenges for the interaction design. For portrait photos, interesting features are centered around the subject itself, that is, people are interested in finding photos with interesting facial expressions, gestures, or poses. For landscape photos, other features are often more important, e.g., focus, view point, zoom setting, and color, whereas the scene itself is static.

While our interface is aimed at the general public, we would like to extend our work and focus on a more specialized group in the future, namely professional photographers, who often take even more similar photos, e.g., at weddings, studio photo sessions, photo-shoots for a magazine, etc. It would be interesting to further tailor our framework to match their needs.

Finally, while interaction will remain essential to the task of photo triaging, there are a number of opportunities for additional automation and intelligence, which can further reduce the load on the user and minimize the necessary time-consuming interactions. We would like to explore the application of computer vision and computer graphics techniques to enhance photowork activities as well as to integrate photo composition capabilities.

APPENDIX: INTERACTIVE SEGMENTATION

Interactive image segmentation has been around for a while in computer vision and computer graphics [18,21,25]. Our contribution here is that corresponding image elements are segmented out from multiple images simultaneously. The user only needs to mark the element(s) of interest on one of the photos, which is a novel feature to our knowledge.

The segmentation engine in our system adopts the general framework of graphcut-based segmentation [7, 18, 25]. The goal is to label each pixel x as either foreground ($l_x = 1$) or background ($l_x = 0$). For this labeling problem, we only consider a local region in the image, specifically a bounding box that surrounds the brush strokes. This is to reduce the computation time, which is crucial for making our interactive system responsive on a tablet computer. Note that we also down-sample the images for faster computation.

Given a local region to work with, a probabilistic color model for the foreground is first computed using the color information of the brushed pixels. It is represented by a Gaussian Mixture Model (GMM) with 5 components ($p^f(\cdot)$). At the same time, we construct a GMM for the background ($p^b(\cdot)$) by randomly sampling 10% of the pixels in the region that was not brushed. Then, the optimized labels for all pixels in the image are computed by minimizing the following energy term using the graphcut solver from [8,30]:

$$E = E_d + \omega E_s \, , \qquad (1)$$

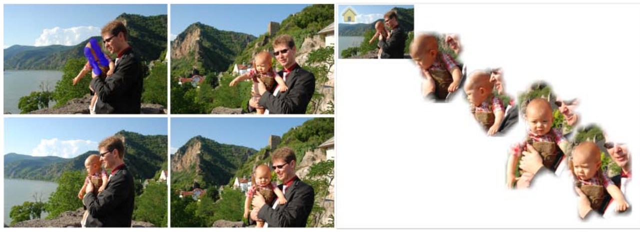

Figure 10. Interactive segmentation with brushing: note the variety of baby poses as well as the camera motion; our engine can still reasonably segment them out at interactive speed.

where E_d is the data term:

$$E_d(l_x) = \begin{cases} K & \forall x \in S \\ l_x L^f + (1 - l_x) L^b & \forall x \notin S \end{cases}, \quad (2)$$

$L^f = -\ln p^f(I_x)$, $L^b = -\ln p^b(I_x)$ with I_x being the color at pixel x, S is the set of brushed pixels, and K is a constant. The smoothness is enforced between adjacent pixels ($\mathbf{p}$,$\mathbf{q}$) with the weight ω and the following cost function [25]:

$$E_s(l_\mathbf{p}, l_\mathbf{q}) = \begin{cases} 0 & \text{if} \quad l_\mathbf{p} = l_\mathbf{q} \\ e^{-\beta ||\mathbf{I_p} - \mathbf{I_q}||^2} & \text{if} \quad l_\mathbf{p} \neq l_\mathbf{q}. \end{cases} \quad (3)$$

where $\beta = (2 < ||\mathbf{I_p} - \mathbf{I_q}||^2 >)^{-1}$ and $< . >$ denotes the expectation over an image sample. For details, please see [25] and references therein. Note that since the image element selected by the user as well as the image background have similar color profiles throughout the set of images, we are able to reuse the GMMs which were trained using the pixels selected by the user in the base image for segmenting all the other images instead of requiring the user to brush the image element in each of the image, hence saving the user the extra work and time needed to train the GMMs for each image.

Figure 10 shows the segmentation results for a case with a fair amount of camera motion among photos. Since the position of the subject is relatively similar, our segmentation algorithm still performs well enough for our task of comparing the brushed elements, even though the results may not be good enough for other applications.

ACKNOWLEDGMENTS

This work is supported by a research grant from the Human Sixth Sense Program at ADSC from Singapore's Agency for Science, Technology and Research (A*STAR), as well as A*STAR SERC grant no. 092 101 0063.

REFERENCES

1. Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., and Cohen, M. Interactive digital photomontage. *ACM Transactions on Graphics (Proc. SIGGRAPH) 23*, 3 (2004), 294–302.
2. Ames, M., Eckles, D., Naaman, M., Spasojevic, M., and House, N. Requirements for mobile photoware. *Journal of Personal and Ubiquitous Computing 14* (2010), 95–109.
3. Apted, T., Kay, J., and Quigley, A. Tabletop sharing of digital photographs for the elderly. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)* (2006), 781–790.
4. Ballan, L., Brostow, G. J., Puwein, J., and Pollefeys, M. Unstructured video-based rendering: Interactive exploration of casually captured videos. *ACM Transactions on Graphics (Proc. SIGGRAPH) 29*, 4 (2010), 87:1–87:11.
5. Bederson, B. PhotoMesa: A zoomable image browser using quantum treemaps and bubblemaps. In *Proc. Annual ACM Symposium on User Interface Software and Technology* (2001), 71–80.
6. Böhmer, M., and Bauer, G. Exploiting the icon arrangement on mobile devices as information source for context-awareness. In *Proc. 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)* (2010), 195–198.
7. Boykov, Y., and Jolly, M.-P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proc. International Conference on Computer Vision (ICCV)* (2001), 105–112.
8. Boykov, Y., and Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 9 (2004), 1124–1137.
9. Cui, J., Wen, F., Xiao, R., Tian, Y., and Tang, X. EasyAlbum: An interactive photo annotation system based on face clustering and re-ranking. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)* (2007), 367–376.
10. Durrant, A., Rowland, D., Kirk, D. S., Benford, S., Fischer, J. E., and McAuley, D. Automics: Souvenir generating photoware for theme parks. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)* (2011), 1767–1776.
11. Frohlich, D., Kuchinsky, A., Pering, C., Don, A., and Ariss, S. Requirements for photoware. In *Proc. ACM Conference on Computer Supported Cooperative Work* (2002), 166–175.
12. Graham, A., Garcia-Molina, H., Paepcke, A., and Winograd, T. Time as essence for photo browsing through personal digital libraries. In *Proc. ACM/IEEE-CS Joint Conference on Digital Libraries* (2002), 326–335.
13. Henze, N., and Boll, S. Evaluation of an off-screen visualization for magic lens and dynamic peephole interfaces. In *Proc. 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)* (2010), 191–194.
14. Huynh, D. F., Drucker, S. M., Baudisch, P., and Wong, C. Time quilt: Scaling up zoomable photo browsers for large, unstructured photo collections. In *Proc. CHI Extended Abstracts* (2005), 1937–1940.

15. Kang, H., and Shneiderman, B. Visualization methods for personal photo collections: Browsing and searching in the PhotoFinder. In *Proc. IEEE International Conference on Multimedia and Expo (ICME)* (2000), 1539 – 1542.

16. Kirk, D., Sellen, A., Rother, C., and Wood, K. Understanding photowork. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)* (2006), 761–770.

17. Kratz, S., Brodien, I., and Rohs, M. Semi-automatic zooming for mobile map navigation. In *Proc. 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)* (2010), 63–72.

18. Liu, J., Sun, J., and Shum, H.-Y. Paint selection. *ACM Transactions on Graphics (Proc. SIGGRAPH) 28*, 3 (2009), 69:1–69:7.

19. McAdam, C., Pinkerton, C., and Brewster, S. A. Novel interfaces for digital cameras and camera phones. In *Proc. 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)* (2010), 143–152.

20. Miller, A. D., and Edwards, W. K. Give and take: a study of consumer photo-sharing culture and practice. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)* (2007), 347–356.

21. Mortensen, E. N., and Barrett, W. A. Intelligent scissors for image composition. In *Proc. SIGGRAPH* (1995), 191–198.

22. Mulloni, A., Dünser, A., and Schmalstieg, D. Zooming interfaces for augmented reality browsers. In *Proc. 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)* (2010), 161–170.

23. Platt, J. C., Czerwinski, M., and Field, B. A. PhotoTOC: Automatic clustering for browsing personal photographs. Tech. Rep. MSR-TR-2002-17, Microsoft Research, 2002.

24. Rodden, K., and Wood, K. R. How do people manage their digital photographs? In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)* (2003), 409–416.

25. Rother, C., Kolmogorov, V., and Blake, A. ”GrabCut”: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (Proc. SIGGRAPH) 23*, 3 (2004), 309–314.

26. Savakis, A. E., Etz, S. P., and Loui, A. C. Evaluation of image appeal in consumer photography. In *Proc. SPIE Human Vision and Electronic Imaging*, vol. 3959 (San Jose, CA, Jan. 23–28, 2000), 111–120.

27. Shneiderman, B., Bederson, B. B., and Drucker, S. M. Find that photo! Interface strategies to annotate, browse, and share. *Communications of the ACM 49* (April 2006), 69–71.

28. Shneiderman, B., and Kang, H. Direct annotation: A drag-and-drop strategy for labeling photos. In *Proc. International Conference on Information Visualization* (2000), 88–95.

29. Snavely, N., Seitz, S. M., and Szeliski, R. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics (Proc. SIGGRAPH) 25*, 3 (2006), 835–846.

30. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. A comparative study of energy minimization methods for Markov random fields. In *Proc. European Conference on Computer Vision (ECCV)* (2006), 16–29.

31. Van House, N. A. Collocated photo sharing, story-telling, and the performance of self. *International Journal of Human-Computer Studies 67*, 12 (2009), 1073–1086.