INSERT: Efficient Sorting of Images on Mobile Devices

Roman Ganhör
Multidisciplinary Design Group
TU Wien, Austria
roman.ganhoer@tuwien.ac.at

Florian Güldenpfennig
Human Computer Interaction Group
TU Wien, Austria
florian.gueldenpfennig@tuwien.ac.at

ABSTRACT

We amass increasing amounts of photos on our mobile devices, primarily captured by built-in cameras. These cameras provide precious opportunities to preserve memories or serve for creative engagement. However, creating order over these vast photo collections gets more difficult as we create more and more photos and this puts these valuable resources at risk. People fail to sort their photo collections manually and automated algorithms are not yet able to identify and group images based on the features that are most relevant to the human beholder. For these reasons we present INSERT, a novel mobile phone application for supporting manual sorting of photo collections in an efficient fashion. A user study featuring 21 participants showed that the proposed interaction mechanisms were well perceived and that there is yet much research to be conducted aiming at the management of image collections on mobile device, in particular with small screens.

Author Keywords

Mobile phones; photography; image management; file management; Design, Experimentations, Human Factors.

ACM Classification Keywords

H.5.2 [**Information Interfaces and Presentation (e.g. HCI)**]: User Interfaces – *Graphical user interfaces (GUI), input devices and strategies, interaction styles, screen design*

INTRODUCTION

While the numbers of photos captured with mobile devices grows rapidly, people usually do not sort or manage their assets (Whittaker et al., 2010). However, operating systems like Android or iOS offer automatized grouping based on location or time (Elliott, 2014; Ybanez, 2014). In addition, there is intensive research to advance sophisticated algorithms in pattern recognition to pool images, e.g., photos depicting the same persons (Darwaish et al., 2014). Even though automatized grouping can be a powerful and convenient feature, it often does not satisfy the users' intentions or needs. All too often the discrepancy between user expectation and automated results lead to frustrating user experiences. Thus, it comes as little surprise that there are also less 'mechanical' and much more playful approaches to exploring photo collections, e.g., as presented by Ott et al., 2012.

As an example for a task that is hard to automate, consider a user compiling a photo book as a keepsake. The most *precious* photos for such a personal compilation cannot easily be determined by a computer algorithm as this involves questions of personal preference, relevance, and the wealth of memories. Algorithms may satisfyingly recognize photographical standards like exposure time or even classify different objects and persons, but can these features make a photo book, which tells the story that we want to be told?

Another relevant example would be an architect, who wants to group all photos made during the process of planning and building a house. These photos will probably depict diverse content, such as sketches made during discussions, small-scale models of the house, various states of the construction progress and finally the completed house. However, the photos just mentioned do not share common attributes such as date, place or content. Thus, automated algorithms will probably fail when grouping the *right* photos for a specific architectural project.

Switching from still photography to video editing, storytelling also fulfills an important and integral part in the searching and sorting process. In video editing it is not only video grouping that matters, it is also the order of the footage that is important for narrating a story (Barrett, 2008). Again, algorithms exist in products (Muvee, 2015) and are presented in academic research (Zsombori et al, 2011) that are mimicking the human ability of storytelling. However, these algorithms will hardly be able to obtain the desired artistic goal as expected from films or videos and thus, automatized video editing is mainly employed by 'leisure artists' to avoid the burden of video editing.

Research focusing on manual and automated photo sorting and grouping, as well as work on managing videos on desktop computers, laptop computers or even table computers is intensively discussed in academia (Hilliges et al., 2007), and a plethora of differing applications exist (Laurie, 2014). Nevertheless, research and applications for touch based handheld devices is still rare in this area, dealing mostly with novel alternatives to browsing large multimedia collections (Dragicevic et al., 2008; Sun et al., 2008) or particularly dealing with the affordances of small screens (Patel et al., 2004). While

OzCHI '15 , December 07 - 10 2015, Melbourne, VIC, Australia
Copyright © 2015 ACM 978-1-4503-3673-4/15/12... $15.00
http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx.

browsing is an important task, its outcome is very volatile without the possibility to store the browsing results. Storing browsing results can depict useful information as it groups digital assets according to an overall user objective.

As the quality of mobile devices in terms of cameras and displays is ever increasing and as technology like cloud computing allows users ubiquitous access to their media assets, we face new challenges. Mobile devices become a common and preferred means to present, receive and discuss multimedia data. Thus, nowadays users carry out tasks in the mobile context they were not able to do before, such as manually sorting and grouping digital assets.

In this paper, we evaluate existing research in the field of mobile multimedia and combine several aspects for sorting and grouping multimedia assets on mobile handheld devices in our proposed interface. The final interface, INSERT, is designed for *fluent* and *efficient* use and was qualitatively evaluated with 21 participants depicting its potential strengths and weaknesses. In more detail, we state the research problem of this paper as follows.

Research Statement

Considering the ever-increasing amount of media assets made with and stored on mobile devices, the importance of these assets' order for the purpose of storytelling and the lack of academic research were the main motives for a closer examination of this research area. Hereby a main challenge for feasible and effective interfaces on mobile devices is leveraging their possibilities while evading their constraints (Xiao et al., 2010), e.g., balancing the image (*thumbnail*) size in dependency to the screen size. Here, the opposing goals are having big thumbnails for good content perception versus as many thumbnails as possible on the screen at once for a good overview.

As indicated by the set of different application examples from above (photo book, documentation of architecture, editing and searching in video), we don't target a specific application area. That is, in this paper, we are not so much interested in, say, supporting specifically creating photo books or photo documenting. Rather, we seek to operate on a more *abstract* level to investigate and design for the elementary task of the searching for and sorting of images on mobile devices (be it personal photos, video frames, etc.). Thus, the objective of this paper is to evaluate the basic interaction mechanisms as proposed by INSERT, independent from too specific application domains. This qualitative feedback again we intend to use for further design iterations of this software. We go on to motivate INSERT drawing on related background literature.

BACKGROUND

Smartphone interfaces vary in many qualities from desktop interfaces due to various evident differences such screen size or input modalities. However, smartphones are increasingly becoming more powerful and users are starting to carry out even complex tasks on their mobile devices. Research in the area of multimedia applications stretches out in various directions: examining the interaction possibilities and affordances of touch-based devices or investigating and extending the boundaries of the limiting factors.

Affordances of Touch Based Devices

In contrast to desktop computers mobile devices are not stationary and hence, can be brought to different contexts. Modern sensor technology enriches mobile applications and enables novel aspects in interaction design. Mobile devices can potentially be used anytime and anywhere to fulfill various kinds of tasks. Users can employ their fingers for 'direct manipulation', i.e., there is no mapping necessary between hand and cursor. The user's finger marks the spot where the input and reaction take place, i.e., the most direct feedback is delivered to the user.

Limiting Factors

Despite the unquestioned possibilities smartphones offer, they suffer from some inherent limitations. One of the most apparent advantages turns out being a major disadvantage with respect to interface and interaction design - *the size*. Both the human eye and screen display have a limited resolution and interface elements must feature a specific size to be (easily) perceivable. Since the size of the smartphone itself is restricted compared to, e.g., a LCD display, the number of elements per screen is limited as well. Due to this constraint even simple interactions often have to be distributed across several screens. Hence, maximizing screen elements, avoiding clutter and providing a smooth user experience is an area of conflict for designers and researchers (Gong et al., 2004).

Thumbnails generally represent video clips and pictures, depicting a small copy of the original content. Studies about the size of thumbnails show that users can identify the content on even comparatively small thumbnails (Hürst et al., 2011). However, while thumbnails are useful for identifying the general content of an asset, they are not appropriate for identifying detailed differences between assets. On one hand, the smaller the thumbnails are the more content can be displayed at once. On the other hand, the bigger the thumbnails are the better the details of one thumbnail can be identified.

Keyboard and mouse are among the most common input devices for desktop computers. The strength of the mouse is its resolution and accuracy, besides, the mouse pointer does not occlude elements on the screen. Experienced users often use keyboard shortcuts as a very efficient input method when interacting with a computer. Touch based devices in contrast have to deal with more inaccurate input methods (Forlines et al., 2007), as the thumb cannot reliably be mapped to a pixel or even a small pixel area. The limiting factor is the size of a fingertip. Additionally, it is difficult to implement shortcuts for touch-based devices without restricting other interaction metaphors.

In contrast to most stationary computer systems, mobile devices allow handling in two orientations, horizontally and vertically. However, the 'natural' orientation for smartphones is vertically as the device can be grabbed firmly and effortlessly with one hand. Depending on the size of the smartphone it can even be operated with one hand. Albeit the 'natural' orientation mobile applications can sometimes be more effective when used horizontally. Besides rather trivial and obvious observations of existing application exist little guidelines that can be generalized for novel interaction mechanisms. Therefore it is up the designer to argue for a novel application's preferred orientation.

Sorting and Rearranging Images on Mobile Devices

Even though grouping thumbnails on mobile devices became a feature widely adopted exploiting meta-data, manually sorting, rearranging, and refining is still not supported by the standards apps of the major mobile platforms. Third party applications have to be purchased or free software has to be obtained to allow the user to rearrange their media items according to their personal preferences. *F-Stop Media Gallery* for Android, for example, allows the user to pick up a media item with a long press gesture, move it to the favored position and drop it wherever wanted.

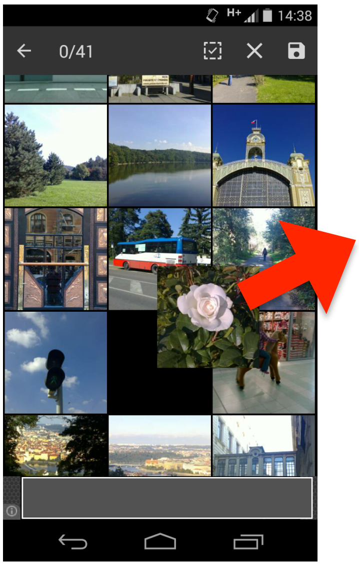

Figure 1: Rearranging photos with a standard grid layout. The rose-picture is dragged into the direction of the arrow.

This process is illustrated by Figure 1, which depicts the interface while inserting a media item (featuring a the picture of a rose) between two other media items. While this approach is "straight forward" and complies with the guidelines for touched based devices it still has a few drawbacks.

First, moving to the start or to the end in a long list of items can be tedious. A series of quick swipes can overcome this, however, this gesture cannot be done when moving a media item at the same time (i.e., fast browsing and drag&drop "don't mix"). The item would drop right after the first swipe when the finger leaves the surface preparing for the second swipe. Second, the interface does not provide an overview over the total collection of media items. This is especially true since the latest updates in the style guidelines of mobile interface manufacturers (e.g., for Android) omit the permanent presence of a scrollbar[1]. Third, the interface does not provide the process of creating new collections from compilations of media items sharing some aspects important for the user. In the common image viewer applications, dragging & dropping an image will result in moving this particular image to another location, not creating a copy. That is, certain pictures can only exist in one collection/folder at a time, and an additional file manager application is needed for instantiating a duplicate. In reality, however, it is a different picture. A photo of an ancient house taken during holidays, for example, can potentially be an interesting member to two different collections: *Holiday in France* and *Houses*. To overcome the first two limitations discussed above various approaches were proposed, some well known are described in the following subsection.

Adapting for Mobile Devices

Focus+Context is an established desktop computer approach for combining the inspection of details (large thumbnails) with the investigation of broader contextual information (small thumbnails) (Furnas, 1986; Rao et al., 1994) in one (distorted) photo mosaic. This approach is also named as *Fisheye View* as it reminds of a *fisheye lens* where just the center of an image is displayed proportionally correct while the outer bounds of the image are visually distorted. *Focus+Context* was also implemented for mobile devices, dating back to 1997 where some interaction features where built on PDA's (Holmquist, 1997). More recently, an application was created for the iPad, which projects thumbnails onto spherical objects (Ahlström et al., 2012) in a *Focus+Context* fashion (see Figure 2).

Overview+Detail is another approach dealing with the challenge of representing data on different levels of detail. The software divides the screen into two distinct areas. One area shows an overview over all or most part of the available information, whereas the second area displays details of a specific subset of information. Thus, an *Overview+Detail* interface design is characterized by the simultaneous display of both an overview and detailed view of an information space. Even though its interface is separated into two distinctive views, user interaction within one view is reflected in the other immediately (e.g., manipulations on the detailed level will be visible on the overview level).

The *Dominant Color Diagram* (Schöffmann et al, 2010) can be described as a combination of *Focus+Context* and *Overview+Detail*. It takes the concepts of *context* and *overview* and extend them to an 'extreme' level whenever necessary. The authors of the algorithm implemented an example where a television show was rendered by means of the *Dominant Color Diagram*, and each frame of the television show is represented by a one-pixel bar in the diagram. While the one pixel diagram does not convey

[1] http://developer.android.com/design/index.html

much information, it provides a valuable overview of the complete data set.

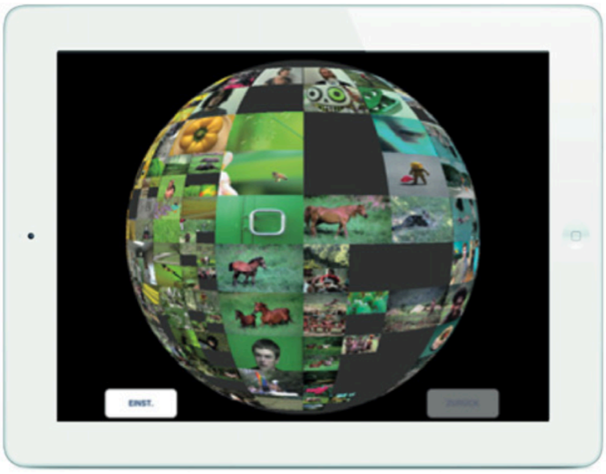

Figure 2: Spherical Projection - mobile example of *Focus+Context* (Ahlström et al., 2012).

Many variations and combinations of the aforementioned approaches exist for both, desktop and mobile computers. Tailoring the most appropriate interactions and combinations to a specific application depicts an important challenge for designers and engineers.

IMPLEMENTATION OF INSERT

The design and implementation of INSERT was motivated by overcoming the limitations of current implementations as described in the section above. This comprises in particular the tedious task of scrolling through a vast amount of media assets and the absence of contextual information. The interface of INSERT and its interaction design particularly considers the strengths and weaknesses of mobile devices and also builds on successful design elements as outlined in the previous section. Novel features of INSERT comprise the implicit definition of distinctive collections when ordering media items and the possibility to add one media asset to several collections in a (and this is one of our hypotheses) efficient and appropriate fashion.

In more detail, INSERT is a user interface for sorting and grouping media assets on touch-based smartphones. The application focuses on three tasks for sorting and grouping media assets that we identified as important in the literature and in our own research: *browsing*, *selecting*, and *filing*.

Browsing through photo collections with a detailed view while at the same time providing contextual information leverages the idea of *Focus+Context*. Thus, the interface provides a visualization of all photos whereas it also conveys detailed information about at least one specific photo at the same time. The interaction concept also supports fast browsing (context) when skimming through the complete collection and precise browsing (focus) when identifying a specific asset (see Figure 3 first and second row).

Selecting a single photo for sorting or grouping is the next step in the interaction cascade of INSERT. We carefully designed the interaction of selecting a media item in an unambiguous/precise fashion to deliver an enjoyable and productive user experience.

Filing photos completes the application as proposed by INSERT and refers to sorting and grouping. This interaction consists of two steps, dragging a photo from the original collection and dropping it to a *selection* (see Figure 3 for an illustration of a *selection*). Thereby, dropping a photo to a *selection* defines its position in this *selection*. In contrast to the drag and drop interaction metaphor, INSERT treats the original collection as 'read only'. This is motivated by two considerations. First, users generally have a good knowledge of their original collections, knowing their structure and thus, removing photos from the original collection could potentially lead to irritation. Second, leaving an item in its original place allows the user to add an item to more than one collection.

Additionally, three non-interactive design features were considered to optimize the user experience. First, most of the screen estate was used by the program leaving little space unused. Second, the interface was kept plain and simple without the need for additional menus or similar complex application structures. Third, sorting photos and compiling groups of photos were considered equally important tasks.

Interface

INSERT was implemented for Android mobile phones. The development device featured operating system version 4.4 and a screen resolution of 1280x720 pixels. The application is optimized for landscape mode for practical reasons. This orientation allowed us to make best use of the screen estate (e.g., in adjusting the overview bar) and, moreover, we are also interested in using INSERT for video assets where the default orientation is widescreen. It distinguishes two main areas (see Figure 3). One area allows users to browse their original collection (browsing) and a second area allows organizing (i.e., filing) the media items in individual *selections* (selecting).

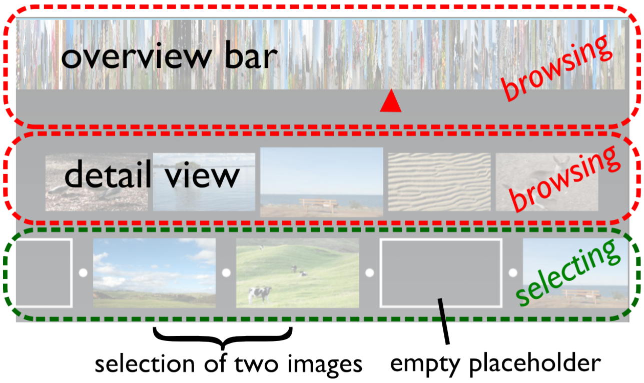

Figure 3: Different areas of the INSERT interface.

The browsing area is divided into two parts, one depicting an overview and the other exhibiting details. In the overview all photos of the original collection are 'squeezed' into the width of the device. This feature was inspired, among others, by work from information visualization (Tang et al., 2009; Viegas et al., 2004). While this operation distorts the photos, it provides the user with rich contextual information. Underneath the overview section an arrow indicates the current position within the data collection, defining the focus or the area of interest of the user. The detail area shows five thumbnails as extracted from the photo subset marked by the arrow.

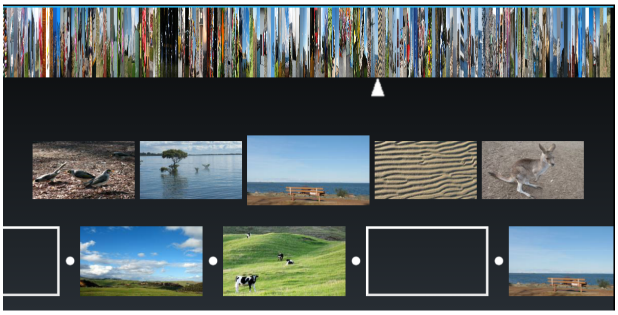

Figure 4: Screenshot of the implementation of INSERT.

Figure 4 shows a screenshot of the actual implementation of INSERT depicting the threefold division as described above. The upper third is filled with small bars, each bar representing a compressed picture. The second third contains five thumbnails big enough to perceive the content of the associating photo. The thumbnail in the center is slightly bigger then the thumbnails to the left and right and is associated with the position of the arrow. The last third comprises a slider holding the photos ordered and grouped in separate collections (*selections*). While the upper two thirds always depict the complete original collection, the lower third (slider) only presents a small fraction of the individual *selections*. The slider can be moved to the left and to the right. In Figure 4 the slider depicts a small *selection* holding two photos, one with a landscape and fluffy clouds on it and one showing green hills with cows. An empty slot defines the end of a *selection* respectively the beginning of the next *selection*.

Interaction

The interaction mechanisms consist of swipes for browsing, drag and drop for selecting and filing. Figure 5 illustrates the swipe and Figure 6 the drag and drop interaction.

Browsing: the task has two levels, coarse and fine. Coarse browsing denotes jumping to an arbitrary point within the thumbnail collection quickly (Figure 5, a), whereas fine browsing is to find a particular thumbnail within a narrow area (Figure 5, b). Browsing the *selections* is executed by swiping over the *selection* area (Figure 5, c).

For selecting and filing the thumbnails into individual *selections* the user drags the corresponding thumbnail from the detail area and drops it onto the *selection* area (Figure 6, d and e). A long click on any of the thumbnails marks a thumbnail for dragging and lifting the finger drops the thumbnail. If a thumbnail is dropped on a regular placeholder (Figure 6, d) that placeholder is filled with the thumbnail. If the placeholder already holds a thumbnail the existing thumbnail will be overwritten with the newer thumbnail. Thumbnails can also be dropped between placeholders (Figure 6, e), the area marked with a bullet. Whenever a thumbnail is dropped on a bullet, the bullet turns into a regular placeholder populated with the thumbnail and all thumbnails right of the new placeholder move one position to the right. This is described in Figure 6 when the selection order of w,x,y,z turns into w,x,y,m,z. To remove a thumbnail from a placeholder the thumbnail gets dragged and dropped as depicted in Figure 6f.

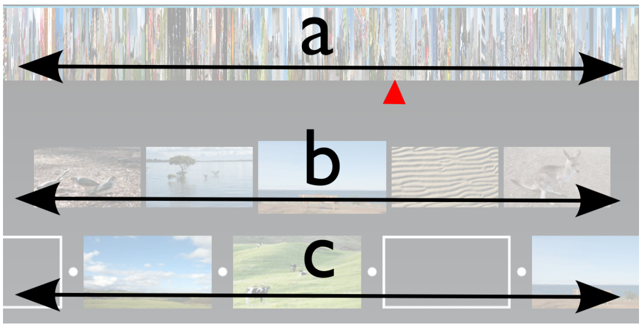

Figure 5: Interactions areas (a,b,c) for swipes gestures.

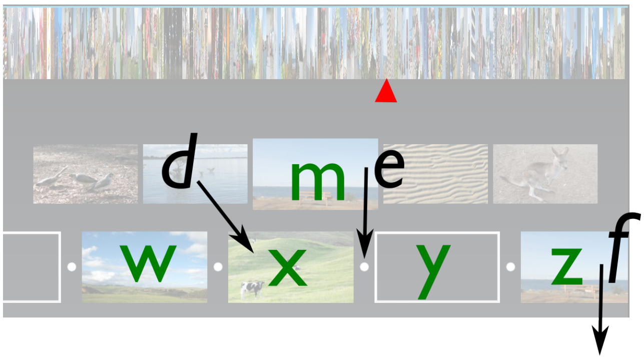

Figure 6: Interactions areas (d,e,f) for drag&drop gestures.

We suggest that the strength of this application is given by the simplicity of the interaction mechanisms while at the same time providing a powerful tool allowing the users to conveniently arrange and re-arrange all photos stored on their smartphones. Since only thumbnails in the *selections* can be deleted, the original set of thumbnails is not altered. INSERT aids re-arranging the order of thumbnails in a media collection and the pooling of thumbnails in *selections*. Thereby, two empty placeholders, one leading and one trailing a set of thumbnails automatically define a selection.

EVALUATION

The design of the interface and the interaction mechanisms were built on prior work as presented above and we followed common design guidelines, and best practices in building the system. Nevertheless, INSERT depicts a novel design concept and consequently we were keen on gathering user feedback on the interface and the interaction. We decided for a qualitative evaluation approach to gain *holistic* insights about this new mechanism for sorting images on smartphones and recruited 21 participants in the age from 16 to 45 through our extended social network. Evaluations sessions were video taped and the participants' comments were transcribed for later analysis.

Study Setup / Procedure

Each evaluation session was initialized with a short questionnaire about the creation and consumption of media data on mobile devices, especially photos. Furthermore, we surveyed their habits regarding mobile phones and how they make use of the capabilities of their devices. Then the participants were explained a use-case tailored to their personal habits, interests and capabilities that involved the re-organization of their media assets. The use-cases were formed on the fly and asked as a question or presented as an argument, like: *Wouldn't it be nice to have a best-of collection of your kid seeing it growing?* or *With the right app you easily could make various collections of different styles of graffiti*. The use-cases were meant to create interest in an application that can solve their brand-new demands. Subsequently the participants received instructions for INSERT and its user interface and were invited to 'play with' 150 prepared pictures. While doing so they were encouraged repeatedly to speak-out loud about their spontaneous thoughts. The whole process, including questionnaire and familiarizing, lasted six to thirteen minutes. As the interface of INSERT waives rich affordances (it is designed for users with some experience and to function as an *efficient* sorting tool) the participants needed an explanation of its interaction mechanics. The length of this training period depended on the participants' familiarity with such tools or mobile phone apps in general. However, after the initial training all participants' were able to utilize the interface sufficiently.

After the participants felt comfortable with the interface an elaborated exercise was given. The purpose of this exercise was to provide a set of real life picture-sorting tasks involving all features offered by INSERT. As such, the participants were asked to identify all pictures containing a red car, and insert them into a new photo *selection*. Subsequently, they had to move particular images (photos that also contained human beings) of the *selection* to yet another *selection*, and eventually delete some of these pictures (containing cars by a specific manufacturer). Again, the participants were encouraged to speak out loud their thoughts, positive or negative. The comments of the participants were recorded and used for later analysis (see next section). The exercises lasted additional 10-15 minutes and subsequently the participants were asked about their concluding opinion about INSERT and for suggestions for further improvements.

Analysis

For analysis, we used an adapted approach to thematic analysis as described by Braun and Clarke (2006). That is, we transcribed the recorded user tests and iteratively coded this data to let prominent themes emerge. Hence, this was an inductive process of coding. Still, it also involved deduction as we were interested in a specific set of a priori categories (and we also framed our questions to the participants accordingly), for example, the *mobile creation and consumption behavior of the participants* or the *overall user experience* (see next section). As stated above, other themes or patterns emerged inductively; in particular, in the discussion section we draw on such exemplary salient user comments that we identified as *important* during analysis. *Importance* or *salience* is defined by the interpretation and judgment of the authors, i.e. comments or patterns don't have to occur often to be relevant.

Study Results

Mobile Creation and Consumption Behavior of the Participants

All participants owned a smartphone with photo capability and used their device on a regular basis for general purposes. Nine participants captured at least five photos per week (*power users*). Seven participants utilized the camera of their smartphone at least once a week (*casual users*) and five participants used their smartphone camera less than once a week (*irregular users*). The average age of the three user groups is depicted in Table 1.

	Mean	Mean Dev.
power-user	28,44	5,72
casual user	29,28	8,04
irregular user	34,2	8,64

Table 1: Mean age and mean deviation for user groups

Sorting and Grouping Habits of the Participants

When asked for their photo managing efforts only 5 of the 9 power users (5/21 total) stated that they did some kind of sorting, all other participants denied grouping images etc. However, after the evaluation session each participant was asked for their interest in using software like INSERT for photo management. 15 out of 21 now indicated that they wanted to receive and use a private copy of the software (see Table 2).

	yes	no
power-user	8	1
casual user	4	3
irregular user	3	2

Table 2: Number of participants interested in an app like INSERT for grouping and sorting (after the evaluation session)

Immutable Original Collections

Based on the presumption that users know the approximate structure of their photo collection, the authors decided to leave the original collection immutable, i.e., not removing pictures or changing their order. Even in the relatively short familiarization period (for the 150 images) of our study it was evident that the users quickly gained a ‘feeling’ for the photo structure of the unaltered photo collection. When asked about their judgment on this design decision, 10 out of 21 participants thought it was very useful (see Table 3). Despites this rather mixed feedback (at least on the first glance), *immutable collections* probed valuable feedback and was heavily commented. This will be discussed in the next section.

	yes	no
power-user	5	4
casual user	3	4
irregular user	2	3

Table 3: Number of persons preferring for immutable original collections.

Overall User Experience

After the ‘hands-on’ evaluation the participants were asked about the overall user experience they had. This question also probed a lot of comments and resulted in rather polarizing opinions (*rather good* versus *rather bad*) as summed up in Table 4. We go on to discuss this finding in the following section.

	rather good	*rather bad*
power-user	7	2
casual user	4	3
irregular user	3	2

Table 4: Final judgment of INSERT’s user experience

DISCUSSION

In this section the authors discuss their design decisions against the outcome of the evaluation. For the qualitative evaluation small exercises were designed and the participants were encouraged to think-out aloud (see above). Consecutively, we asked for ideas, changes or improvements regarding the interface and interaction design concept. These comments of the participants were pooled and assigned to a design decision. By doing this, the authors collected statements favoring or refusing single design decisions.

The initial idea for INSERT was to provide a useful and easy to use tool for arranging photos on mobile devices. Providing overview while supporting precise interaction mechanism was another important concept. Allowing the sorting and grouping of *all* photos on one *single* screen (i.e., not spanning several screens on the mobile phone) was maybe the single most important and controversially discussed design choice. In contrast, allowing the user to put a media item in more than one collection was a rather unanimous decision. In the following paragraphs we discuss comments, notes and supplementary ideas we found relevant. They are represented by salient comments we identified during analysis.

"The app should mark the items [in the original collection] that are already in one of the selections." (P3) Most participants, even those, who did not prefer immutable original collections, proposed to *mark items* in the original collection. Proposals ranged from using a colored frame around an item and highlighting its usage in a *selection* to placing small numbers in a corner of the item indicating the number of the associated *selection*. This improvement could aid in identifying specific photos in larger *selections*.

Design decisions were always made in favor for an effective and fluent interaction. Even obvious drawbacks of a decision were taken into account to push efficiency. One such decision was allowing thumbnails in the selections to get overridden by another thumbnails without any request. The underlying motivation of this was that users could easily get annoyed when rearranging an existing selection. Not surprisingly, several participants made comments like *"Oh, that [thumbnail in the selection] gets overridden very easily. Maybe this should be confirmed?" (P7)* However, when the authors discussed this issue with the participants many of them agreed that recurring confirmations would be annoying. A reasonable compromise could be an *undo function*.

The *selections* as displayed in the last row in the interface are separated by one empty placeholder (Figure 6, marked as 'y'). When scrolling through *selections* the termination placeholder can easily be missed and the user (unknowingly) navigates to the next *selection*. P19 as well as several other participants noted that *"... this stripe on the bottom holding the selection […] should stop automatically whenever a new selection starts."* (P19) The drawback of such an automatic stop after each *selection* could result in tedious re-swiping to go from one *selection* to the next. To overcome this several approaches were discussed. A promising idea proposed by one participant was a combination of simple gestures.

"Always a long click ... why do drag and drop always need a long click?" (P4) Filing a thumbnail from the original collection to the *selections* was completed by a drag and drop gesture. We initially decided to stick to the original Android metaphor for dragging a screen item (long click on the screen item). As it turned out the participants accustomed to the workflow very quickly and felt hampered by the comparable time consuming and non-productive long click. A few participants (P1, P4, P13) highlighted that the swipe interaction for fine browsing is horizontal, whereas adding a thumbnail to a *selection* requires a vertical gesture. Thus, these gestures will hardly interfere mutually. However, an unfamiliar

implementation of a familiar gesture could potentially lead to confusion on the users' side.

As users interacted with INSERT they quickly got used to its capabilities. Six users (five of them power-users) tried to probe its limits and created dozens of *selections*, each containing dozens of thumbnails. When sliding left and right in the *selection* area users can easily get confused about their overall position and their position in a *selection*. To overcome this drawback various proposals were made. One proposal recommended a pair of sliders indicating the overall positions and the position in the *selection*. Another proposal was put into the words: *"[m]inimize the selection and represent it with just one [thumbnail]" (P11).* The metaphor for minimizing could be implemented by one double click. Whenever a double click is carried out on a thumbnail in a *selection* the *selection* collapses into one thumbnail. This one thumbnail is marked to identify it as a representative for a whole *selection* (e.g., by a red border frame). A double click on such a representative thumbnail again would expand a collapsed *selection*. However, the double click gesture was also proposed for maximizing a thumbnail to full screen mode or for naming a *selection*.

"This [...] would come in handy for video editing" (P7) These considerations regarding the handling of *selections* (see previous paragraph) led to some interesting reflections around possible application domains that go beyond simple image sorting. As one of the participants, a professional video editor, mentioned, he was regretting the lack of feasible video editing software for mobile devices. In his opinion, and after some further investigations into this matter we do agree, that the easy way to duplicate and insert photos of INSERT, might also be valuable for mobile video editing apps. In this domain, it is often required to create multiple copies of short video clips or clip fragments to be further processed and *inserted* into the target movie. Thus, for future work, it may well be worth the time investigating whether mechanisms of INSERT can be used to create video editors for the mobile devices.

"But what will happen if the overview bar is too small for my photo collection?" (P6) P6 realized an important limitation of the current implementation of INSERT when she pointed out that the overview bar wouldn't be able to contain an infinite number of images. While this fact is true, this limitation can be avoided by implementing established UI ideas and concepts. An adapted version of *Focus+Context* can increase the amount of photos displayed, i.e. instead of *every* photo every second (third, fourth, etc.) photo is represented with one bar. To indicate these omissions between images, the corresponding bar is indicated by a reduced height (i.e., bars representing more images are shorter). Since the user can still browse through the photos one by one this is no considerable drawback for the usability, and the interface still provides a contextual overview.

Broader Reflections

The partly passionate discussion with the participants during our evaluation showed their strong interest and the potential for advanced multimedia tasks on smartphones. As we kept the initial design simple and concentrated on the basic tasks for sorting and grouping several questions arose about future extensions. Participants also asked for the integration of INSERT into other mobile application, expanding their functionality. This however touches upon one of the essential questions in mobile application design: should we concentrate on one single task like sorting/grouping or should we provide additional features, as well? The study results from this paper indicate that for the purpose of image sorting the first option should be favored, even though this was not always verbalized by the participants. Future work is needed for answering this question of complexity more reliably.

FUTURE WORK

Choosing a qualitative interview and user narrative approach when investigating INSERT allowed us to understand the way participants experienced the application. As discussed before, compressing all available images into one overview bar (see Figure 3) turned out to be an effective and appealing feature, if not the most interesting feature according to the participants. For this reason, we plan to further investigate this interaction mechanism employing *quantitative* methods to complement our qualitative data. Thus, we are currently implementing an alternative application identical to INSERT, named *InsertGrid*, however, with a conventional horizontal scrollable grid of photos instead of the overview bar. Hence, *InsertGrid* only shows a small subset of all available photos simultaneously. We will compare INSERT against *InsertGrid* and evaluate a set of benchmarks such as time to complete and total errors, allowing us to describe the efficiency of the overview bar statistically while keeping all other variables constant. We favor evaluating isolated UI elements of INSERT systematically (e.g., the overview bar; the single interaction steps) over comparing INSERT with an existing image sorting applications (“gold standard”) as it yields more insights about the various elements of INSERT. Additionally, investigating various variables at the same time (i.e., comparing INSERT with a gold standard) exacerbates identifying and explaining causalities. In parallel, we intend to advance our application with respect to two concrete steps.

First, we started implementing some of the participants’ comments as reported above into a new version of the application. Second, in this iterated version we also integrated online logging capabilities, i.e., user interactions are sent to our interaction log server. We plan to deploy INSERT to Google’s app store to gather a larger data set of participants who will use the proposed interface on their own devices. This additional data will complement our evaluation with quantitative observations and external validity. A final issue will be the question whether or not the concept of INSERT is able to scale to very large collections of photos and what adjustments to the interface and interaction design have to be made to support the sorting of many images.

REFERENCES

Ahlström, D., Hudelist, M. A., Schoeffmann, K., & Schaefer, G. 2012. A user study on image browsing on touchscreens. In Proceedings of the 20th ACM international conference on Multimedia (pp. 925-928). ACM.

Barrett, H. 2008. Digital storytelling. University of Alaska Anchorage. Retrieved February 1, 2015.

Braun, V ., & Clarke, V . 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2, 77-101.

Darwaish, S.F., Moradian, E., Rahmani, T., Knauer, M. Biometric Identification on Android Smartphones, Procedia Computer Science, Volume 35, 2014, Pages 832-841

Dragicevic, P., Ramos, G., Bibliowitcz, J., Nowrouzezahrai, D., Balakrishnan, R., Singh, K. 2008. Video browsing by direct manipulation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 237-246). ACM.

Elliott, M. 2014. How and where to find your photos in iOS 8. cnet.com. http://www.cnet.com/how-to/how-and-where-to-find-your-photos-in-ios-8/

Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R. 2007. Direct-touch vs. mouse input for tabletop displays. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 647-656). ACM.

Furnas, G. W. 1986. Generalized fisheye views. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '86), Marilyn Mantei and Peter Orbeton (Eds.). ACM

Gong, J., Tarasewich, P. 2004. Guidelines for handheld mobile device interface design. In Proceedings of DSI 2004 Annual Meeting (pp. 3751-3756).

Hilliges, O., Baur, D., and Butz, A. 2007. Photohelix: Browsing, sorting and sharing digital photo collections. In Horizontal Interactive Human-Computer Systems. TABLETOP'07. Second Annual IEEE International Workshop on (pp. 87-94).

Holmquist, L. E. (1997, March). Focus+ context visualization with flip zooming and the zoom browser. In CHI'97 Extended Abstracts on Human Factors in Computing Systems (pp. 263-264). ACM.

Hürst, W., Snoek, C.G.M., Spoel, W-J. and Tomin, M. 2011. Size matters! how thumbnail number, size, and motion influence mobile video retrieval. In Proceedings of the 17th international conference on Advances in multimedia modeling - Volume Part II (MMM'11)

Laurie, V. 2014. Best Free Digital Photo Organizer. Gizmo's Freeware. http://www.techsupportalert.com/best-free-digital-photo-organizer.htm

Muvee Technologies Pte. Ltd. 2015. Muvee. http://www.muvee.com/products/muvee-reveal-11

Ott, C., Hebecker, R., & Wakes, S. 2012, Picture the space: three concepts for management and presentation of personal digital photographs. In *Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction* (pp. 1-8). ACM.

Patel, D., Marsden, G., Jones, S., & Jones, M. 2004. An evaluation of techniques for browsing photograph collections on small displays. In *Mobile Human-Computer Interaction-MobileHCI 2004* (pp. 132-143). Springer Berlin Heidelberg.

Rao, R., Card, S. K. 1994. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 318-322). ACM.

Schoeffmann, K., Taschwer, M. and Boeszoermenyi, L. 2010. The video explorer: a tool for navigation and searching within a single video based on fast content analysis. In Proceedings of the first annual ACM SIGMM conference on Multimedia systems (MMSys '10). ACM, p247-258.

Sun, Q., Hurst, W. (2008). Video Browsing on Handheld Devices - Interface Designs for the Next Generation of Mobile Video Players. MultiMedia, IEEE, 15(3), 76-83.

Tang, Anthony, Saul Greenberg, and Sidney Fels. "Exploring video streams using slit-tear visualizations." CHI'09 Extended Abstracts on Human Factors in Computing Systems. ACM, 2009.

Viégas, Fernanda B., Martin Wattenberg, and Kushal Dave. "Studying cooperation and conflict between authors with history flow visualizations." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2004.

Whittaker, S., Bergman, O., & Clough, P. (2010). Easy on that trigger dad: a study of long term family photo retrieval. Personal and Ubiquitous Computing, 14(1), 31-43.

Xiao, J., Lyons, N., Atkins, C. B., Gao, Y., Chao, H., and Zhang, X. 2010. iPhotobook: creating photo books on mobile devices. In Proceedings of the international conference on Multimedia (pp. 1551-1554). ACM.

Ybanez, A. 2014. Impala for Android recognizes your pics and automatically sorts them into categories. Android Authority. http://www.androidauthority.com/impala-android-366554/

Zsombori, V., Frantzis, M., Guimaraes, R. L., Ursu, M.F., Cesar, C., Kegel, I., Craigie R. and Bulterman, D.C.A. 2011. Automatic generation of video narratives from shared UGC. In Proceedings of the 22nd ACM conference on Hypertext and hypermedia. ACM.