Combining Multitouch Gestures and Sketches to Explore Photo Collections

Paulo Gonçalves
Dept. of Computer Science and Engineering
INESC-ID/IST/TU Lisbon, Portugal
Email: paulo.goncalves@tecnico.ulisboa.pt

***Abstract*—In this paper we present a novel solution for interactive exploration of photo collections. This approach takes advantage of a seamless combination of multitouch and sketch-based gestures. Based on previous work and user research, including participatory design sessions, on the photography field, we successfully tested and validated such take on photo exploration. The obtained results and feedback led us to believe that sketch-based interaction might well be the missing piece on touch interaction, a solution to provide touch applications with meaningful shortcuts.**

I. Introduction

Digital media formats have a tremendous impact in people's life. Photography can be seen as one of the most improved areas with the transition to a digital format. Such transition promoted a serious decrease in the costs and the spreading of cameras across different devices. Moreover, the evolution of social and photography related networks are also a driving force in the change of photo shooting habits and the most relevant for the last few years.

These factors contribute to a huge increase of people's personal photo collection's size. As collections grow in size people tend to neglect organization [1], ending up with a huge amount of unorganised data. This would not be a problem if there was a written base for the collection as text-based algorithms that aim for organization are much more efficient. Photo-tagging would be a reasonable solution but it is very time-consuming and highly subjective. In the absence of manual organization there is still the information collected by the camera and the photos characteristics. Many solutions have been developed based on this assumption but there is still no efficient and pleasant way to browse large collections.

Motivated by the aforesaid premises and inspired by the work already done on this matter, a new approach to the problem will be taken. This work will tackle the problem by enhancing user's interaction with the system. Therefore, rather than empowering the system by trying to make it mimic human's thinking, power will be given to who already has it, the user. As a result of this approach, a novel interaction technique for exploration of photo collections, developed for multitouch devices, will be presented.

Although, the real aim of this work is to explore the possibilities created by the use of a sketched-based interaction. Consequently, the subject of photo exploration is just the scenario used to test the introduction of sketch-based interfaces in multitouch applications. However, multitouch interfaces have already been consumed by multitouch gestures that are now deeply settled as the standard form of interaction for such devices. Therefore, this work does not only intend to assess the use of sketch-based interaction but also to determine if it is possible for this type of interaction to coexist with the current standard multitouch gestures, in a seamless fashion.

A close inspection to the work done in the field of photo exploration was performed. It was important to fully comprehend the proposed scenario so it would not hinder the testing of this work's main goal. After that first probe, a survey and a set of participatory design sessions were conducted in order to make clear who would the target users are and to puzzle out their needs and behaviour.

It was then necessary to establish how multitouch and sketch-based gestures would be combined and each one's purpose. During the research phase, it was possible to unveil two distinct kinds of tasks regarding photo exploration: to browse through the collection and to search upon it. Therefore, multitouch gestures were employed to allow browsing throughout the collection, allowing the user to directly manipulate the presented visualization, while the sketch-based gestures were used to let users invoke filters from a provided set, in order to be able to perform search tasks through filtering.

The aforementioned solution was implemented in a functional prototype and put to test by eleven users. To assess the validity and usability of the sketch-based interaction, a toolbar menu was added to the prototype. That way, it was possible to draw a comparison between a fully multitouch approach and the combined approach. Testing results were encouraging, as sketch-based gestures collected the preference of users during a comparison testing. Users found out that this approach was more direct and made the experience more fun and enjoyable.

The good testing results and feedback gathered from users are a guarantee that sketch-based gestures are a viable form of interaction for touch devices. Moreover, this type of interaction might well fill one important gap that still exists in touch interfaces, the absence of applications' shortcuts. The proposed gestures proved to be ideal as a simpler and more direct interaction than regular menus. Sketch-based gestures might well be the missing link on touchscreen interfaces, providing both regular and power users with meaningful shortcuts.

II. Related Work

Several works done in the fields of photo exploration and visualization, that bear some resemblance to the purpose of this work, were firstly analysed. The goal was to find what has already been done on this matter, as a basis and inspiration for the development of this work.

Time Quilt [2] emphasises the importance of the temporal distribution of a photo collection based on the principle that people most commonly want to browse their photos by event than by querying them based on more specific properties. This work was developed taking into account not only the assumption of the importance of such temporal characteristics but also an efficient spacial distribution, trying to minimize the white space between photos. *Huynh et al.* perfectly demonstrated people's tendency to browse photos by their time frame. This aspect was not neglected and was taken into account as a fundamental influence for the development of the ongoing work.

Strong, Hoeber and Gong [3] proposed a system based on a 2D virtual desktop to perform visualization and exploration of large image collections. The solution is based on the concept of image discovery, when the mental concept of the desired image is far vague and incomplete. Vibe's browsing interface is based on pan and dynamic zoom operations, adjusting the number of images shown according to the level of zoom applied and available space in the screen. This solution showed to be an improvement over the commonly known grid layout in terms of search results perception and ease of use. Vibe proved to be a clean exploration technique that when applied to a multi-touch interface might lead to even more promising results. Vibe's browsing approach was seen as the biggest inspiration for that particular aspect of this work.

FreeEye's project tried to implement a shift on the way interactive image browsing systems are designed. Instead of focusing on learning the way users create related semantics between the pictures, *Ren and Kalic* [4] propose a system more user-centered designed augmenting user's interaction with more content, facilitating a more intuitive and effortless browsing. The only way possible to interact with the interface is through clicking on a image, triggering a reorganization of the interface where the clicked image will now be the center image. While trying to be extremely innovative, this solution has several problems that seriously compromise its use. The driving force used on the development, the definition of intuition, showed to be vague as it could not be proved and maybe cannot even be measured. On the practical side, the interface is hugely limited by the effort put on facilitating the interaction by only allowing to click images. Moreover, the inability to show an overview of the full collection is a tremendous handicap when dealing with large amounts of pictures.

Hilliges et al [5] claim that any photo software should try to support at least a specific set of typical activities [1] performed over media collections: filing, selecting, sharing and browsing. Along with this features the system should offer flexible interactions for browsing and sharing images. Tabletops are seen as the ideal device to implement such a system as they allow to mimic the flexibility and tangibility of physical media. To achieve the aforementioned requirements some design goals were identified: overview at all times, details on demand, support for temporary structures and flexible spatial arrangements. The collection visualization on Photohelix is based on a spiral representing a timeline, where photos are organized according to the capture date. Although the system was designed and implemented only to assess the validity of requirements and design goals, not to work under realistic circumstances, the authors could verify the importance of a time-based exploration in browsing a personal collection. Also, the possibility of rearranging interface elements, either by reallocating controls or freely move the photos, proved to help improving the flexibility of the interface and to close the emotional gap between digital and printed photos handling.

PhotoMagnets [6] is based on the concept of exploratory search, combining querying and browsing strategies to improve user's ability to efficiently explore a photo collection. A flexible interface is the main requirement as *Chen et al.* claim that applications with pre-defined structures cannot support the flexibility observed in user behavior. To achieve a flexible interface three characteristics were considered mandatory: encourage exploration, facilitate serendipity and keep the user in control. In other words, support exploration beyond query-based search, encourage users to discover serendipitous findings while being able to recover from those distractions and prefer interaction and feedback over automation. As preliminary tests showed that chronological and loose representations have advantages, respectively, in supporting structured and unstructured browsing, PhotoMagnets ended up adopting two different solutions. The *TreeBrowser*, aimed for structured browsing, and the *BubbleBrowser*, flexible browsing and searching. To facilitate search's flexibility, the *BubbleBrowser* relies on the magnet metaphor where multiple search criteria act as magnets, attracting matching photos while repelling others. The magnet metaphor proved to be, after user evaluation, especially appealing in facilitating search, clustering and comparison. Magnets helped query formulation and let users understand underlying relations and discover semantic relationships. Although feedback about this approach was highly positive, it deeply relies on tagging quantity and quality which in most of the cases is inexistent. User evaluation also helped to state the importance of time information when exploring a personal collection, as it was one of the most asked features for future development.

Interactive coffee table for exploration of personal photos and videos [7] by *Pedrosa et al.* focused on understanding how interactive tabletops can be used for exploring and viewing home media, such as photos, in family gatherings. In order to achieve such understanding, the authors had to develop a functional prototype that allowed the exploration of photos and videos in a collaborative manner. The developed prototype was created on top of one of the main challenges considered by *Pedrosa et al.*: the desirability of having a digital interface

that uses the metaphor of physical photographs placed on a table. Despite the fact the authors were more enthralled by the collaborative nature of the prototype, there are some useful observations regarding the collection's visualization and exploration. Physical photos' metaphor was considered fun and intuitive, clearly indicating that users approve it.

There are some characteristics that clearly define the type of system previously presented. The most relevant features must be: i) the ability of clearly see an overview of the whole collection, ii) the ability of inspecting a photo in detail, iii) the ability of browse through the collection, iv) the ability to search for a photo, v) the ability to apply filters in order to perform the two previously referred activities and vi) the use of both meta and content-based information. Not all of these characteristics are nor need to be satisfied by each single system, bearing in mind its own specificity. Although, the solution to be presented as a result of this work should satisfy all of them, aiming at being a more complete and flexible system.

III. User Research

As the main stakeholder, the user of any future system needs to be probed and taken into great consideration. Getting to know who the target users are should be among the very first things to accomplish when starting to create an interactive solution, prior to any design or development. More than just knowing who they are, it is also important to be able to acquire a deep understanding on target users characteristics, habits and expectations, so their full satisfaction with the final proposal can be somehow guaranteed.

Evidently, when speaking about a photo exploration tool, it is fairly trivial to assume that the target users will be, somehow, photographers. Although, this definition is far too broad and inaccurate as there are many characteristics that can differentiate various types of photographers. In an attempt to narrow down the users range of this work, it was assumed from the very beginning that its aim would be on non-professional photographers. That said, the target users will be people that casually photograph or that have photography as a hobby.

Starting from the concept of non-professional photographers, split in both casual and enthusiast photographers, and using a simple survey, it was possible to establish distinct characterisations for each one. Although both can be seen as photographers, casual and enthusiast photographers present almost entirely different profiles. Enthusiasts, as expected, proved to be much more dedicated and hardworking towards their photographic activity. While casual photographers just want to record moments, enthusiasts strive to get the best photo possible, making use of all their knowledge and equipment. It is those differences that make them fall apart that will be the biggest challenge for both the design and development of a solution that will suit both. Even though, enthusiasts can be seen as a natural evolution of casual photographers. Comparably, the proposed solution should suit casual photographers' needs and provide some extra tools for more advanced operations that will definitely please the enthusiasts.

IV. Task Research

Although getting to know the target users is essential, it is equally important to thoroughly understand the tasks they currently perform and the ones they would like to perform in a near future. More than identifying what people do and wish they could do, it is also key to comprehend the habits and behaviours behind those tasks.

In order to better understand such habits, two different kinds of workshop were conducted. The first workshop was solely focused on peoples photo collection organisation habits and corresponding mental model. On the other hand, the second tried to anticipate what people would like to be able to accomplish in a near future, in a participatory design session.

The first series of workshops held was called Photo Sorting. This particular workshop was purely designed to be used in this case, under the particular domain of photo browsing and exploration. Inspired by the well-known Card Sorting activity, Photo Sorting relied on the use of actual pictures instead of the usual cards with concepts.

Photo Sorting aimed at giving a better understanding of how people structure and organize their personal photo collections. It focused on the reasoning behind photo grouping and importance prioritisation.

The last workshop was based on PICTIVE, that stands for Plastic Interface for Collaborative Technology Initiative through Video Exploration. It is a well-known participatory design method used to engage users to help developing new graphical user interfaces [8]. Making use of paper mock-ups, PICTIVE gives the user a sense of what a system or a piece of software will look like and how it will behave once it is finished, enabling a non-technical person to contribute with ideas to the development process.

For this work, the design of the interface was not the sole objective of the PICTIVE session as users insight was also important to design the multi-touch interaction. This aspect is particularly important as multi-touch allows more creative and flexible ways of interaction that can still be widely explored. Moreover, the possibility of using ten fingers as input opens even more possibilities. Here, participants' creativity can play an important role to explore those new possibilities.

Summing up, these sessions had three well defined objectives: to help designing the system's graphical interface; to help designing the interaction between the user and the system and, once more, assess how the users browse through their collections.

During the workshops, some interesting actions were observed and important feedback was collected. All those observations and feedback were analysed and used as guidelines for the design of the proposed solution. Time is by far the most important feature when organising and disposing a set of photos and its combination with place, of what can be called an event, carries great meaning. Although it does not have the same degree of importance, people present in the photos is a source of information that should not be neglected. In contrast, interfaces and interaction created during

the participatory design part of the workshop were, for most of the participants, too similar to a point-and-click desktop application. Furthermore, when asked to explore the multi-touch possibilites, participants tended to overdo and imagine excessively complex interactions that added virtually no value to the interaction.

V. Developed Solution

Empowering the user to better perform exploratory tasks upon photo collections is the main premise for this work. Each photo present in the collection contains a lot of information, attached and in the photo itself, that tend to be neglected. All this information has the ability to help the user better understand his own collection and to potentiate the creation of powerful tools to help him better explore it. Although making good use of that information and providing new powerful tools is important, it is also vital that the user feels comfortable using them and that in no way they hinder his tasks. On that account, the main contribution of this work will lie heavily on the way the user interacts with those tools and information.

The main contribution of this work is a novel interaction language that combines both multitouch and sketch-based gestures. This language aims to put the user in total control of the exploratory tasks, providing tools that allow the user to make the most of all information contained in his collection and each photo of it.

Although, the presented solution reflects not only the interaction aspect but all the system that is needed to employ that particular type of interaction. Figure 1 sums up the software architecture behind the solution and will serve as a basis to further exploration of its components.

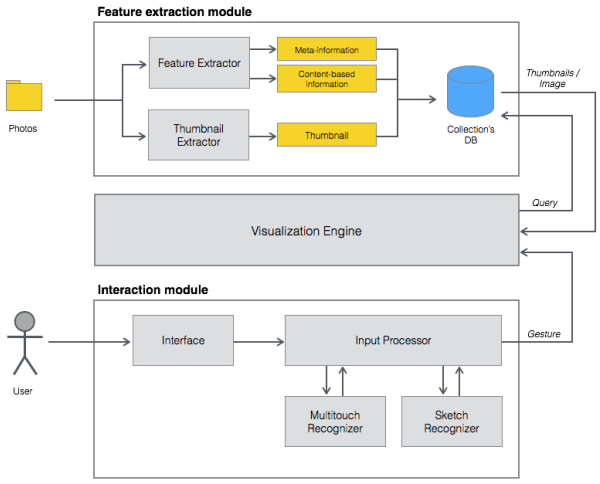


Fig. 1: Architecture Overview

As we can see from Figure 1, the implemented architecture is composed by three layers. However, the three layers can be seen as two main components, the Feature Extraction and the Interaction modules, and a bridge between them, the Visualization Engine. The Interaction module comprises the main front-end portion of the solution as it is responsible to gather and process user's interaction. On the other end, the Feature Extraction module contains the back-end part and it is responsible for storing all information concerning the user's photo collection. The Visualization Engine, the intermediate component of the solution, has to be accounted as part of the front-end side of it, as it is in control of generating the photo collection's visualization that is then shown to the user. However, this component does not only generate the visualization but it is also responsible for the communication between the two main modules. These three layers will now be detailed from top to bottom.

Given a set of photos, a photo collection, the Feature Extraction module will gather every single photo to be processed. Each of these photos will then be handled by two separated extractors, the feature's and the thumbnail's. The Feature Extractor plays the main role in this stage as it is in charge of gathering all the useful information regarding each photo. As a matter of fact, this component can also be decoupled in two simpler ones as it is responsible for extracting both meta and content-based information. Running in parallel to the Feature Extractor is the Thumbnail Extractor. This particular extractor creates a small square sample of the original photo, to be later used by the Visualization Engine.

The visualization engine is responsible for generating the photo collection's visualization that the user will see and then interact with. Consequently, what matters most about this component is to understand the nature of the created photo collection's visualization.

During the user research phase, the most referenced characteristic used to both identify a single photo or organize a photo collection was, undoubtedly, time. The date of a photo capture is, in the vast majority of the cases, the initial characteristic that comes to a user's mind when looking for a particular photo or set of photos. From that first information the user will then build a line of thought that will lead him directly to the desired photos or to other defining characteristics, such as place or event.

Therefore, when the time to start designing the photo collection's visualization came, it was obvious that photos would have to be somehow organized by date. Apart from the fact that time was the most popular characteristic, date is also a very powerful concept commonly used as a comparator, allowing people to frame events in terms of time. This use as a comparator also applies to the case of photo exploration. When a user has a deep understanding over a collection, like the case of personal collections, it is fairly easy to understand if the desired photos were taken before or after a given set of photos. In this case, this approach can also be seen as a guide for user's navigation along the collection. For that reason, it is not only important to have the date information present on the visualization but also to keep the photos' chronological order.

When thinking about organising data in a chronological order, the most common thought is the use of a timeline. The concept of timeline is already well-known and provides a logical sequence of events, which perfectly suits this case. For that reason, the photo collection's visualization, pictured

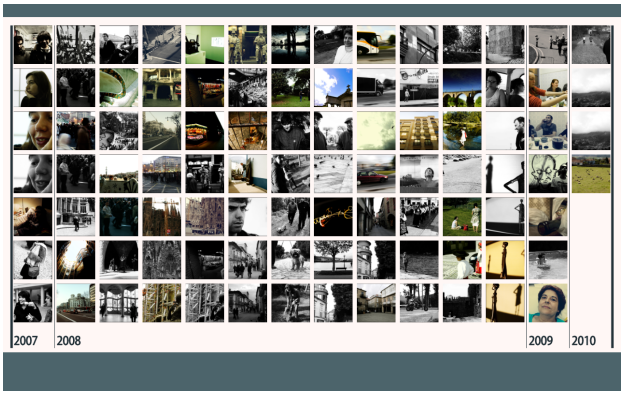
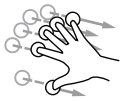


Fig. 2: Photo collection's timeline

TABLE I: Exploration's multitouch gestures

	**Drag three fingers** over a photo set to show more photos of that or previous set.
	**Pull the hands apart** over a photo to increase the time detail. **Pull them together** to decrease detail.
	**Drag one hand** to pan along the collection.

in Figure 2, is based on the timeline principle.

The photo collection's visualization is then divided in multiple sets of photos that have one particular characteristic in common, the time when the photos were captured. Although each set is attached with some time information, this information does not reveal the whole date of capture. Alternatively, there are three different degrees of detail for a set: year, month and day. As all sets do not represent the same degree of detail they cannot be presented at the same level. Therefore, a hierarchical structure is applied to the visualization.

One of the issues that needs to be addressed by the visualization's design is the size of the collections that this work is aimed at. It is reasonable to assume that, in most of the cases, it will not be possible to display the whole collection at a time. Also, other constraint that supports this assumption is the fact that photos' thumbnails have to be big enough so the photos are easily recognisable. In order to successfully address the aforementioned restrictions, it is necessary to apply a summarization technique to keep the sets' size from growing bigger than the screen's boundaries. Consequently, not all photos of a single set might be visible at a given time. The choice of a summarization technique to decide which photos to present would definitely be a question of great discussion. Although, as neither visualization or summarization are the focus of this work and for sake of simplicity, each set is just represented by its first photos.

At first, what is presented to the user is solely the photo collection's visualization. There is a set of multitouch gestures, Table I, that allow the user to manipulate that visualization. As not all photos might be visible at a time, users need a way to retrieve the hidden photos of a particular time set of their collection. By dragging one set of photos, using two or three fingers of one hand, the user is able to unveil more photos. The new photos that are shown might belong to the set that is being dragged, if the hand is moved to the left, or to the previous set, if the hand is moved in the opposite direction. Upon finishing the action, the visualization automatically readjusts to properly display the new photos. As a result of this interaction there will be photos dragged outside the visible screen area. In order to get those photos back in a visible position, there is a pan action that can be performed by dragging a full hand, to the desired direction, over the collection. Retrieving more photos from the collection is one way to explore it in the proposed visualization, other way is to navigate through the time hierarchy of each set. To increase the time detail of one set, for instance to show photos by month in a year set, the user has to pull both hands apart on top of the desired set. This action will cause the visualization to dig in to the next level of detail. In order to revert this action, theres the opposite gesture of pulling hands together over the desired set, then decreasing time detail.

The interaction presented until this point is adequate and sufficient to perform purely exploratory tasks, where the user is not looking for a particular photo or set of photos but rather (re)discovering his collection. Although, as it was already mentioned, in order to achieve the goal of empowering the user to better perform search-based exploratory tasks upon his photo collection, it is necessary to provide him with the appropriate tools. It is important that those tools facilitate such tasks in a way that the changes that they made are completely transparent to the user. To satisfy that premise, all control of those tools must be given to the user, avoiding the use of hidden algorithms that the users have no visibility upon. Summing up, those tools must provide simple operations that the users can easily identify, understand and control. On the other hand and taking into account the huge size of nowadays personal photo collections, it is fairly obvious to presume that the easiest way to facilitate such exploratory tasks is to provide tools that allow the user to reduce the number of photos under inspection. Making good use of the enormous amount of information stored with each photo, it is possible to provide the user with simple operations that rely on that information and that allow the improvement of the accuracy of such tasks. Such tools can rely on more immediate information, like the dominant colour of a photo, or in technical details regarding the photo capture, like shutter speed or aperture. Therefore, the ideal tools to help users perform the aforementioned tasks will be simple filters that can both rely on meta-information or information extracted from the photo content. These tools can also be adapted to both casual and amateur photographers, including or not the more technical aspects of photography.

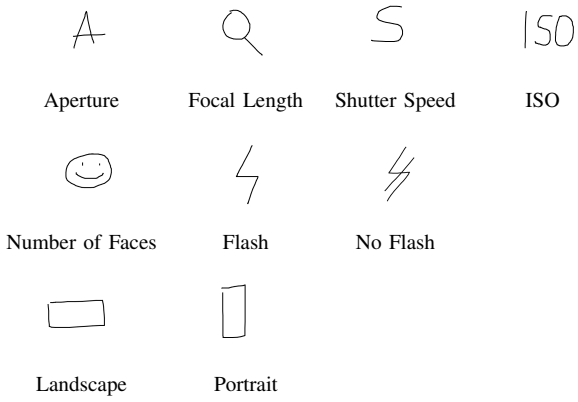
For the sake of simplicity, the use of filters will be described

TABLE II: Available filters

TABLE III: Filters' sketching gesture

Drag one finger to draw a filter's representation and open it in edition mode.

TABLE IV: Filters' invocation sketches



taking into account what was developed for the prototype. A limited amount of information was made available in the developed prototype. Consequently, the number of developed filters is also limited to the available information, EXIF and face detection. Even though, nine filters were implemented to be at the prototype's user's disposal, as depicted in Table II.

As filters are not present on the interface at first, there needs to be a way to access them. The usual and simplest approach to achieve this would be to create a filters menu and associate it with a multitouch gesture. However, that solution is unlikely to be scalable to a greater number of filters and would force the user to go through an extra step to apply a filter, choosing it from the menu. The approach here proposed lets the user skip that step and invoke the filter directly. This direct invocation could be done via multitouch gestures but this particular type of gestures don't allow users to create a relation with the concept they relate to, being extremely difficult to use a multitouch gestures to represent, for instance, a concept like aperture. Multitouch interaction works really well for direct manipulation of a component displayed on the interface, like what happens in the navigation stage of this proposal. In this particular case, it is necessary to use a representation for each filter that the user can easily relate to, being easily recognisable and memorisable. Drawings are something that carry a tremendous expressive power and fill perfectly in the desired description.

Putting everything into context, multitouch is the driver for the photo collection's visualization's manipulation while in order to invoke a filter a sketch input is used. Performing a sketch input is as simple as drawing the pretended sketch with one finger anywhere on the screen, Table III. The presented language allows the usage of both single and multi-stroke sketches, as can be seen by the chosen sketches to represent each filter, Table IV. The associated sketches resemble in some way the meaning of each filter and make use of some already well-known representations of those concepts, so they can be easy to recognize and, more importantly, to remember. After finishing drawing the sketch, the system will evaluate it and will try to recognize it. If it is successfully recognized the filter will appear in the place where the drawing was finished, otherwise an error message will be shown.

It is easy to understand that, apart from the different information sources, those filters have a distinct nature regarding the represented type of information. The first five filters have a numeric nature, having an associated value, while the remaining are purely boolean. This is even more evident by the fact that the boolean filters have, in fact, two distinct representations, illustrating the two opposite states. Having different sketches for each of the possible states allows the user to chose the desired action upon invocation, when drawing the sketch. A similar approach is also valid for the invocation of numeric filters. Although, as these particular filters only present one possible state, the drawing allows the user to give the system a hint about the desired associated value. That hint is given by the size of the sketch drawn by the user, so the bigger it is the bigger the starting value will be. Undeniably, this method does not allow the user to set the desired value immediately but provides an approximation that should facilitate the selection of the final value.

Moreover, there are some actions that are common to all filters, regardless of their nature. Those actions are to apply, to edit and to delete a filter. This actions are pretty straightforward and are based on the existence of three distinct areas on the screen, as can be seen on Figure 3. To apply a filter, after invoking it, the user just needs to drag it to the Active Filters Dock or simply flick it to the bottom of the screen. There, the applied filters will be stored and maintained visible so the user can keep track of the actions that are currently applied to his collection, as depicted in Figure 4. Also, having the applied filters visible allow to quickly pick

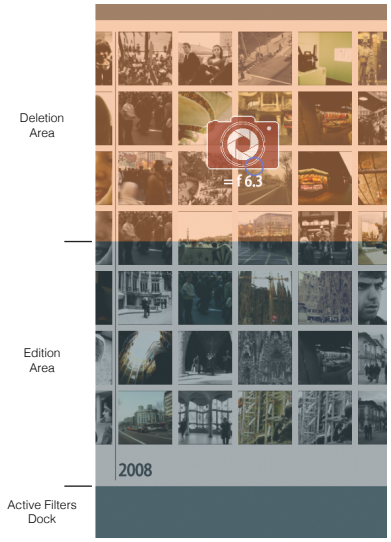


Fig. 3: Filters' manipulation areas

Fig. 4: Active filters dock

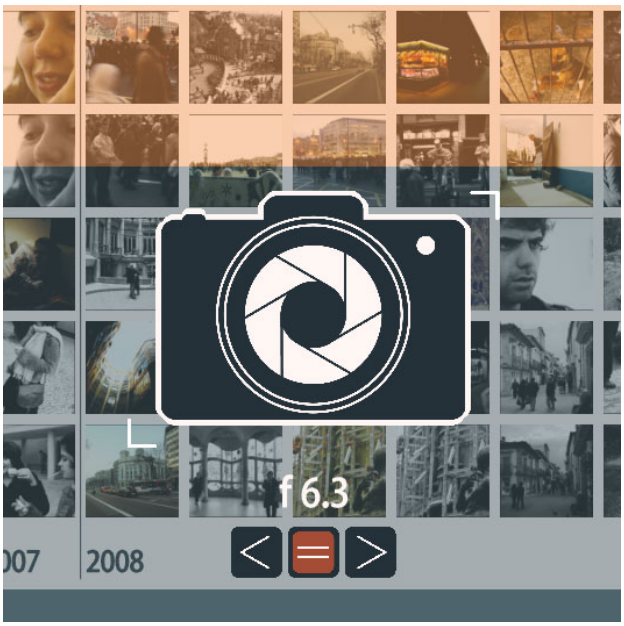


Fig. 5: Aperture filter, a numeric filter

them back to edition. To bring the filters back to edition mode, the user just have to drag them back to the Edition Area. When a user wants to get rid of a particular filter, he just needs to flick it up the screen or drag it to the red Deletion Area on the screen.

Boolean filters' interaction is limited to the aforementioned common actions. On the other hand, numeric filters, as can be seen in Figure 5, need more manipulation. When a numeric filter is invoked the user has the chance to control the initial associated value. Although, this is not sufficient and there needs to be a way to adjust it to the desired value after invocation. In order to configure that value, the user can directly manipulate the filter by stretching it or shrinking it. To perform these actions it is just necessary to, respectively, pull two fingers apart over the filter or just pull them together, just like when handling the full size photo.

The numeric filters still have an extra degree of flexibility. More than setting the associated value, numeric filters also have three different options to help filtering the collection, three comparators: lesser than, equal to and more than the chosen value. In addition to that, it is also possible to combine the presented comparators in order to get even more filtering options: lesser than or equal to and more than or equal to. Numeric filters have another singularity as there is not only one numeric filter of each type but two. Consequently, it is possible to apply two filters from the same type in order to create an interval of desired values. The solution will prevent users to select incorrect combinations of comparators or even between the two filters of the same type.

Furthermore, there is no restriction to the application of multiple filters at the same time, regardless of their nature. Users can freely combine all filters at their disposal until they are satisfied with the amount of visible photos or the target photos are found.

The proposed interface was thought to be flexible enough to support both pure discovery and search-based exploratory tasks. This flexibility comes in great part from the simplistic nature of the interface. At first, the user is presented with nothing more than the visualization of his collection. The ability to directly manipulate the visualization with a set of multitouch gestures concedes enough freedom to users to perform discovery tasks. On the other hand, search-based tasks are supported by the use of the aforementioned filters. Each of those filters has a sketch representation that is use to invoke it. Drawing that representation anywhere on the screen allows the user not only to invoke the respective filter but in some degree to configure it upon invocation, making use of alternative representations or different sketch sizes.

Summing up, the proposed interaction language uses multitouch to directly manipulate the interface elements, such as the visualization and the already invoked filters, and the sketch-based gestures to allow the invocation of filters. Putting everything into context, multitouch gestures allow users to freely explore the collection while sketch-based gestures allow them to filter it. The innovative factor of the proposed interaction language lies in the fact that the two types of gestures integrate in such a seamless way that they are always available and there is no need to change between interaction modes.

VI. User Evaluation

There are several ways of evaluating interactive solutions, ranging from different kinds of expert reviews to user testing. In this case, as the aim was to understand users' response to a new interaction language, the obvious path to follow was user testing.

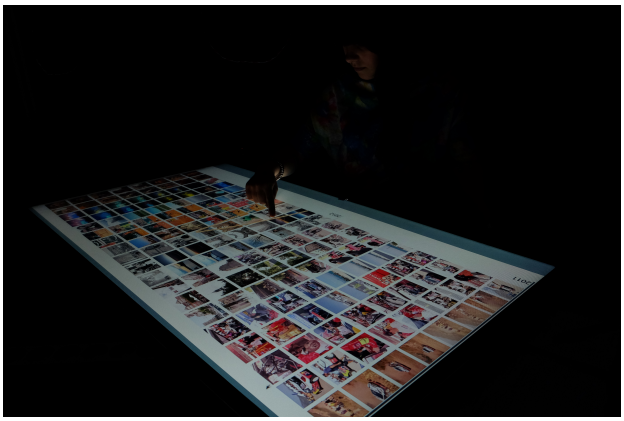


Fig. 6: User performing a task

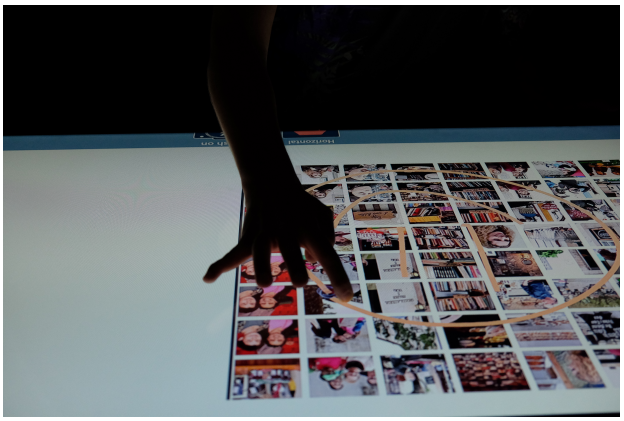


Fig. 7: User using a sketch-based gesture

Although, there are several methods that can be applied when user testing. The first decision was to evaluate what would be the best methods to be used. Having a new interaction language to evaluate there was the need to compare it with something that is already somehow familiar to users and that they don't have any issue understanding. Therefore, there was the need to provide users with a well-known alternative to the sketch-based filter invocation, in order to establish the testing session on top of a comparison testing. This procedure allows users to interact with two distinct ways of achieving the same output, allowing the drawing of a smart comparison between them.

As a result of the aforementioned decision to use a comparison testing, a toolbar menu with the different available filters was introduced in the previously presented solution and implemented on the functional prototype.

Just like during the design phase, the testing stage requires some standards to be met. Like the design workshops, the testing sessions need to be held with representatives of the target user groups that were chosen in the very beginning of this work. Therefore, it was necessary to look for enthusiasts and casual photographers to perform these testing sessions.

In order to evaluate the use of sketch-based gestures, there is firstly the need to evaluate the proposed gestures. The two main attributes to be evaluated are sketches' recognition and memorability. Therefore, the first actual test of the session was to understand if users could recognize the proposed sketches. To achieve this goal, users were asked to try to identify, using their knowledge about photography, the sketches that were represented on Table IV. Enthusiast photographers were asked to identify the whole set of sketches while casual photographers were only asked about the ones they were more familiar with: number of faces; flash; no flash; landscape and portrait.

To actually test the proposed solution, users were asked to perform a set of four tasks, as can be seen in Figure 6, in three different conditions: firstly only using the toolbar menu; secondly only using the sketch-based gesture and finally being able to use the approach that pleased them more. All tasks were then timed and evaluated according to the SEQ (Single Ease Question) method, a 7-point rating scale to assess how difficult users find a task. After completing the first two rounds of four tasks, users evaluated their experience with the prototype, first with the toolbar menu and then with the sketch-based gestures, by answering a SUS (System Usability Scale) questionnaire.

The whole testing procedure was designed as setup for the third, and final, set of tasks where an ultimate comparison testing was performed in order to understand users preference regarding the toolbar or the sketches.

VII. Results and Findings

Eleven user testing sessions, following the aforementioned procedure, were conducted. The user sample included four enthusiasts and seven casual photographers, with ages comprehended between early 20s and mid 30s. All users were familiar with touch-screen interfaces, using them everyday on their smartphones and/or tablets.

Concerning the recognition and memorability rates for the provided sketches, the results were encouraging. Over 80% of recognition and even higher rates of memorability, 100% for casual photographers, can be seen as a sign of a easy and fast learning process towards the use of a sketch-based interaction. Despite that, the memorability tests were performed just after the usage of the system, relying on short-term memory. It would be interesting to observe the evolution of the results when the user has to count on his long-term memory.

In terms of difficulty and starting with the time results it is fair to assume that the collected times should not be used to draw any conclusions. The problems that almost every user faced with the sketches' recognition had a deep impact on the final times. Fortunately, SEQ was used with the same of objective. In this case to prevent the harm caused by the recognition issues, Figure 7, users were asked to give a response that would not take into account any problem faced with the sketches' recognition. According to the obtained results, independently from the type of interaction chosen by users, all tasks proved to be easy to complete, all results over 6, and provided a near tie. Regarding the perceived usability of the two presented approaches there is one more time a

TABLE V: User Testing - testing

User	VE	E	H	VH
U1	Menu	**Sketch**	Menu	Menu
U2	**Sketch**	**Sketch**	**Sketch**	Menu
U3	**Sketch**	**Sketch**	Menu	Menu
U4	**Sketch**	Both	Menu	Both
U5	**Sketch**	**Sketch**	**Sketch**	**Sketch**
U6	**Sketch**	**Sketch**	**Sketch**	**Sketch**
U7	**Sketch**	**Sketch**	**Sketch**	**Sketch**
U8	**Sketch**	**Sketch**	**Sketch**	**Sketch**
U9	**Sketch**	**Sketch**	Menu	**Sketch**
U10	**Sketch**	**Sketch**	Menu	Menu
U11	**Sketch**	**Sketch**	**Sketch**	**Sketch**

draw. The presented results did not allow to come up with any smart verdict. However, from what was observed and from what users said during the execution of the tasks, the main problems they faced came up when editing the filters' values and were not directly related with the toolbar menu or with the drawing of the sketches.

Finally, the comparison testing presented the most interesting results among all the gathered data,. Far from the ties observed on the other tests, the comparison testing revealed the users' preference for the sketch-based interaction. As can be seen on Table V, users widely preferred to use the sketch-based interaction when both approaches were available, being exclusively used in 32 of the 44 performed tasks and combined with the toolbar menu in other two occasions. This preference was even more crystal clear for the easiest tasks, when the toolbar menu was only used one time alone and another one combined with the sketch interaction. Although, for the hardest tasks, the results were once again a near tie, still revealing a slight advantage for the sketch-based interaction.

When asked about their preferences, users were unanimous ranking the sketch-based interaction as their favourite and the most direct. According to the majority of the users, this approach proved to be simpler, in the sense that it would lead to a quicker interaction, having less steps than the use of the toolbar menu. The ability to provide the system with more information by using different sized sketches proved to be crucial for this result. Moreover, users also agreed that the use of sketch-based gestures made the experience of using the system more fun and enjoyable. Although, when there was a need to apply multiple filters at the same time, user conceded that the toolbar would be faster, as in this particular cases would need less steps to perform all actions.

Regarding the memorability and the learning curve for the sketch-based interaction, the feedback was that this kind of interaction was far more intuitive than multitouch gestures or the commonly used keyboard shortcuts. This resulted of the visual nature of sketches that, according to users, facilitate the learning and the mapping of the gestures with their represented concept. One of the users also stated that drawings are a very common form of communication and that they are closest to the human natural communication than the aforementioned alternatives. Moreover, several users referred that visual memory is way more powerful than muscular memory, when comparing the use of sketches to keyboard shortcuts.

VIII. Conclusions and Future Work

The main goal for this work was to propose an interaction language able to combine both multitouch and sketch-based gestures in a seamless way. That said, photo collections' exploration was the chosen scenario to work upon and apply the aforementioned language. Research was made on the selected scenario, through the analysis of the work already done but also through the contact with prospect users. As a result, a new interaction language, combining the two already referred types of gestures, was proposed. The proposed solution was then implemented in a fully functional prototype and subject to user testing. Results and feedback were gathered and presented in the last chapter.

Sketch-based gestures proved to be extremely effective when dealing with simple and direct tasks. Users praised the intuitive and visual nature of sketches and believed they posed a quicker and more immediate alternative to the use of a toolbar menu.

The perception of being a more immediate action, representing a more direct path towards the action the user has in mind, than the commonly used toolbar, makes this type of interaction an ideal candidate to be used as shortcut. Moreover, the proven ability to combine sketch-based gestures in a seamless way with multitouch ones, without the need to resort to a completely different mode, does not add any extra complexity to the interaction.

Regarding the learnability of such gestures, users commented that the sketches' visual nature would vastly help to keep a lower learning curve. This is a strong argument for the use of sketch-based gestures as shortcuts, as they would be easier to learn than the keyboard shortcut that almost everyone use everyday.

This work pointed at proposing and validating a interaction that would seamlessly combine multitouch and sketch-based gestures. It is reasonable to assume that such goal was achieved and the feedback gathered from users after testing the concept can demonstrate it. Over and above that, during the development of the prototype and during the first tests, the proposed sketch-based gestures arose as an unexpected implementation of touch-based shortcuts. The concept of using sketch gestures as shortcuts has to be considered the main contribution of the presented work.

The possibility of using sketch-based gestures as shortcuts just surfaced in a later stage of this work. Therefore as its potential was not accounted until a later point, both the design,

development and testing were not guided towards that goal. The absence of a powerful concept to create shortcuts for touchscreen applications makes future work focused on this subject quite likely to happen.

It would be interesting to evaluate the usability of such approach on smartphone and tablet applications. Also, it would be a great opportunity to explore new ways of inputting the sketch gestures, as the one presented on this work would be likely to collide with most applications' interaction.

Apart from the use of this approach as a way of creating shortcuts, evolving the presented solution and prototype to accommodate user's own gestures would probably be the next logical step. Then, the problems faced with sketch recognition should be finally overcame and an even more exhaustive inspection of the real value of this interaction form could be conducted. Also, the creation and use of personal sketches could be evaluated and compared to the use of standard sketches.

Finally, the sketch-based interaction still has a lot of power to be unveiled. The possibility of combining several sketches and create a truly interactive language seems exceptionally appealing. If the addition of the size as an attribute of a sketch-based input had already offered a tremendous upgrade on its expressive power, the ability to combine different sketches in a single command would exponentially increase that expressiveness.

REFERENCES

[1] D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss, "Requirements for photoware," in *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, ser. CSCW '02. New York, NY, USA: ACM, 2002, pp. 166–175. [Online]. Available: http://doi.acm.org/10.1145/587078.587102

[2] D. F. Huynh, S. M. Drucker, P. Baudisch, and C. Wong, "Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections," in *CHI '05 extended abstracts on Human factors in computing systems*, ser. CHI EA '05. New York, NY, USA: ACM, 2005, pp. 1937–1940. [Online]. Available: http://doi.acm.org/10.1145/1056808.1057061

[3] G. Strong, O. Hoeber, and M. Gong, "Visual image browsing and exploration (vibe): user evaluations of image search tasks," in *Proceedings of the 6th international conference on Active media technology*, ser. AMT'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 424–435. [Online]. Available: http://dl.acm.org/citation.cfm?id=1886192.1886250

[4] K. Ren, R. Sarvas, and J. Calic, "Freeeye: intuitive summarisation of photo collections," in *Proceedings of the 17th ACM international conference on Multimedia*, ser. MM '09. New York, NY, USA: ACM, 2009, pp. 1127–1128. [Online]. Available: http://doi.acm.org/10.1145/1631272.1631531

[5] O. Hilliges, D. Baur, and A. Butz, "A.: Photohelix: browsing, sorting and sharing digital photo collections," in *In: TABLETOP 2007: Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, 2007, pp. 87–94.

[6] Y.-X. Chen, M. Reiter, and A. Butz, "Photomagnets: supporting flexible browsing and searching in photo collections," in *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ser. ICMI-MLMI '10. New York, NY, USA: ACM, 2010, pp. 25:1–25:8. [Online]. Available: http://doi.acm.org/10.1145/1891903.1891936

[7] D. Pedrosa, R. L. Guimarães, M. da Graça Pimentel, D. C. Bulterman, and P. Cesar, "Interactive coffee table for exploration of personal photos and videos," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. ACM, 2013, pp. 967–974.

[8] M. J. Muller, "Pictive - an exploration in participatory design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '91. New York, NY, USA: ACM, 1991, pp. 225–231. [Online]. Available: http://doi.acm.org/10.1145/108844.108896