Evaluation of Adaptive SpringLens – a Multi-focus Interface for Exploring Multimedia Collections

Sebastian Stober, Christian Hentschel & Andreas Nürnberger
Data & Knowledge Engineering Group, Otto-von-Guericke-University Magdeburg, Germany
{sebastian.stober, christian.hentschel, andreas.nuernberger}@ovgu.de

ABSTRACT

Sometimes users of a multimedia retrieval system are not able to explicitly state their information need. They rather want to browse a collection in order to get an overview and to discover interesting content. In previous work, we have presented a novel interface implementing a fish-eye-based approach for browsing high-dimensional multimedia data that has been projected onto display space. The impact of projection errors is alleviated by introducing an adaptive non-linear multi-focus zoom lens. This work describes the evaluation of this approach in a user study where participants are asked to solve an exploratory image retrieval task using the SpringLens interface. As a baseline, the usability of the interface is compared to a common pan-and-zoom-based interface. The results of a survey and the analysis of recorded screencasts and eye tracking data are presented.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*Evaluation/methodology*

INTRODUCTION

Growing collections of multimedia data such as images and music require new approaches for exploring a collection's contents. A lot of research in the field of multimedia information retrieval focuses on queries posed as text, by example (e.g. query by humming and query by visual example) as well as automatic tagging and categorization. These approaches, however, have a major drawback – they require the user to be able to formulate a query which can be difficult when the retrieval goal cannot be clearly defined. Finding photos that nicely outline your latest vacation for a presentation to your friends is such a retrieval goal and underlining the presentation by a suitable background music cannot be done with query by example. In previous work [4, 5] we have developed an interface for exploring image and music collections. An overview of the entire collection is given by displaying few spatially well distributed objects as thumbnails for orientation. The rest of the collection is displayed as points. Multi-dimensional scaling (MDS) is applied to generate the initial distribution of objects on the display. Users can enlarge interesting regions with a fish-eye lens [1] that

NordiCHI '10, 16-OCT-2010, Reykjavik, Iceland

Figure 1. User-interface with an object marked green in primary focus and two objects in secondary focus. (color scheme inverted for print)

allows more thumbnails to be displayed at bigger size. The surrounding space is compacted but not hidden from the user to preserve overview. The MDS – as any other dimensionality reduction technique – introduces "projection errors" in the sense that objects that are very close in high-dimensional feature-space might be projected at large distances from each other and objects that are very dissimilar are placed next to each other respectively. This effect is alleviated by automatically adapting a secondary focus consisting of additional fish-eye lenses in regions containing objects similar to those in primary focus. The resulting distortion brings separated nearest neighbors back together. Figure 1 shows the interface. A more detailed description of the interface and the underlying algorithms are given in previous work [4, 5].

This paper presents a user study for evaluating the usability of the approach: Screencasts of 30 participants solving an exploratory retrieval task were recorded together with eye tracking data (using a Tobii T 60 Eye Tracker) and web cam video streams. This data was used to identify emerging search strategies among all users and to analyzed to what extent the primary and secondary focus was used. Moreover, first-hand impressions of the usability of the interface were gathered by letting the participants say aloud whatever they think, feel or remark as they go about their task (think-aloud protocol). The interface is generic and could be applied on any kind of multimedia data. Representatively, this study was confined to visual data (vacation photos) that can be conceived and compared at a glance and without prior knowledge. (In other scenarios such as music retrieval either a good knowledge about the individual songs or more time for comparison is required.) The prototype used for the evaluation relies on MPEG-7 visual descriptors (EdgeHistogram, ScalableColor and ColorLayout) [3, 2] to compute the visual similarity. Based on empirical evaluation, the maximum distance of the feature distances (taken also from MPEG-7 standard) was selected as distance metric.

USER STUDY

The following questions were addressed in the user study: 1. How does the lens-based user-interface compare in terms of usability to common panning & zooming techniques that are very popular in interfaces using a map metaphor (such as Google Maps)? 2. How much do users actually use the secondary focus or would a common fish-eye distortion (i.e. only the primary focus) be sufficient? 3. What interaction patterns do emerge? 4. What can be improved to further support the user and increase user satisfaction? For question 1, the participants compared our approach with common pan & zoo and additionally a combination of both. The interaction approaches are described in detail below. For questions 2 and 3 we recorded and analyzed, how the participants interacted with the system. Answers to question 4 were collected by asking the users directly for missing functionality.

User-Input Controls

The system that was used for the evaluation supports three input control modes: panning & zooming (P&Z), adaptive SpringLens (SL), and the combination of both. Further, common functions are available irregardless of the control mode.

Panning & Zooming (Baseline)

These are very common interaction techniques that can e.g. be found in programs for geo-data visualization or others that make use of the map metaphor. Panning shifts the displayed region whereas zooming decreases or increases it – without affecting the size of the thumbnails (this can be done separately as described below). Using the keyboard, the user can pan with the cursor keys and zoom in and out with $+$ / $-$. Alternatively, the mouse can be used: Clicking and holding the left button while moving the mouse pans the display. The mouse wheel controls the zoom level. If not the whole collection can be displayed, an overview window indicating the current section is shown in the top left corner, otherwise it is hidden. Clicking into the overview window centers the display around the respective point. Further, the user can drag the section indicator around which also results in panning.

Adaptive SpringLens

Holding the right mouse button, the user can move the fish-eye lens of the primary focus around and enlarge regions of interest. Clicking onto an image (with the right mouse button) centers the primary focus on the image. The magnification factor of the lens can be changed using the mouse wheel by holding the right mouse button. As it can become very tiring to hold the right mouse button while moving the focus around, users can toggle a focus lock mode (return key). In this mode, the user clicks once to start a focus change and a second time to freeze the focus. To indicate that the focus is currently being changed (i.e. mouse movement will affect the focus), an icon showing a magnifying glass is displayed in the lower left corner. The secondary focus is by default always updated instantly when the primary focus changes. This behavior can be disabled resulting only in an update of the secondary focus once the primary focus does not change.

Common Functions

A few functionalities have been added in both interfaces to facilitate interaction. Pressing the A key changes which images are chosen to be displayed as thumbnail – either none, those in a focus region, a few (chosen to be representative and non-overlapping), or all. Thumbnail size can be changed with the $PgUp$ and $PgDn$ keys. Double clicking a thumbnail opens a dialog window with the image at big scale that allows the participant to classify the image to a predefined topic by clicking a corresponding button. As a result, the image is marked with the color representing the topic. The complete collection can be filtered by highlighting all thumbnails classified to one topic. This is done by pressing the key for the respective topic (numeric keys). Highlighting is done by focusing a fish-eye lens on every topic member and thus enlarging the corresponding thumbnails.

Experimental Setup

At the beginning of the experiment, the participants were asked several questions to gather general information about their background. Afterwards, they were presented the four image collections in fixed order. On the first collection, a survey supervisor gave a guided introduction to the interface and the possible user actions. Each participant could spent as much time as needed to get used to the interface. Once, the participant was familiar with the controls, he continued with the other collections for which a retrieval task (described below) had to be solved without the help of the supervisor. At this point, the participants were divided into two groups. The first group used only P&Z on the second collection and only SL on the third one whereas the other group started with SL and then used P&Z. The order of the datasets stayed the same for both groups. (This way, effects caused by the order of the approaches and slightly varying difficulties among the collections are avoided.) The fourth collection could then be explored by using both, P&Z and SL. After the completion of the last task, the participants were asked to assess the usability of the different approaches. Furthermore, feedback was collected pointing out, e.g., missing functionality.

Participants

The survey was conducted with 30 participants – all of them graduate or post-graduate students. Their age was between 19 and 32 years (mean 25.5) and 40% were female. Most of the test persons (70%) were computer science students, with half of them having a background in computer vision or user interface design. 43% of the participants stated that they take photos on a regular basis and 30% use software for archiving and sorting their photo collection. The majority (77%) declared that they are open to new user interface concepts.

Test Collections

Four image collection were used during the study. They were drawn from a personal photo collection of the authors.[1]

[1] The collections and topic annotations are publicly available under the Creative Commons Attribution-Noncommercial-Share Alike

Table 1. Photo collections and topics used.

collection	topics
Melbourne & Victoria	–
Barcelona	Tibidabo, Sagrada Família, Stone Hallway in Park Güell, Beach & Sea, Casa Milà
Japan	Owls, Torii, Paintings, Osaka Aquarium, Traditional Clothing
Western Australia	Lizards, Aboriginal Art, Plants (Macro), Birds, Ningaloo Reef

Each collection comprises 350 images – except the first collection (used for the introduction of the user-interface) which only contains 250 images. All images were scaled down to fit 600x600 pixels. For each of the collections 2 to 4, five non-overlapping topics were chosen and the images annotated accordingly. These annotation served as ground truth and were not shown to the participants. Table 1 shows the topics for each collection.

Retrieval Task

For the collections 2 to 4, the participants had to find five (or more) representative images for each of the topics listed in Table 1. For guidance, handouts were prepared that showed the topics – each one printed in a different color –, an optional brief description and two or three sample images giving an impression what to look for. Images representing a topic had to be marked with the topic's color. It was pointed out that the decision whether an image was representative for a group was solely up to the participant and not judged otherwise. There was no time limit for the task. However, the participants were encouraged to skip to the next collection after approximately five minutes as during this time already enough information would have been collected.

Tweaking the Nearest Neighbor Index

In the original implementation, at most five nearest neighbors are retrieved with the additional constraint that their distance to the query object has to be in the 1-percentile of all distances in the collection. (This avoids returning nearest neighbors that are not really close.) 264 of the 1050 images belonging to collections 2 to 4 have a ground truth topic label. For only 61 of these images, one or more of the five nearest neighbors belonged to the same topic and only in these cases, the secondary focus would have displayed something helpful for the given retrieval task. This let us conclude that the feature descriptors used were not sophisticated enough to capture the visual intra-topic similarity. A lot more work would have been involved to improve the features – but this would have been beyond the scope of the study that aimed to evaluate the user-interface and most specifically the secondary focus which differentiates our approach from the common fish-eye techniques. In order not to have the user evaluate the underlying feature representation and the respective similarity metric, the index was modified for the experiment: Every time, the index was queried with an image with a ground truth annotation, the two most similar images from the respective topics were injected into the returned list of nearest neighbors. This ensured that the secondary focus would contain some relevant images.

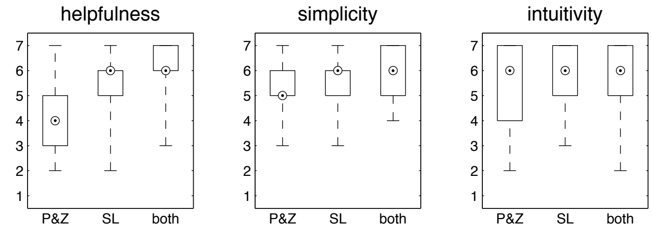

Figure 2. Usability comparison of common panning & zooming (P&Z), adaptive SpringLens (SL) and the combination of both. Ratings were on a 7-point-scale where 7 is best. The box plots show minimum, maximum, median and quartiles for N = 30.

Results

Usability Comparison

Figure 2 shows the results from the survey comparing the usability and helpfulness of the SL approach with baseline P&Z. What becomes immediately evident is that half of the participants rated the SL interface as being significantly more helpful that the simple P&Z interface while being equally complicated in use. The intuitiveness of the SL was surprisingly rated slightly better than for the P&Z interface, which is an interesting outcome since we expected users to be more familiar with P&Z as it is more common in today's user interfaces (e.g. Google Maps). This, however, suggests that interacting with a fish-eye lens can be regarded as intuitive for humans when dealing with large collections. The combination of both got even better ratings but has to be considered noncompetitive here, as it could have had an advantage by always being the last interface used. Participants have had more time for getting used to the handling of the two complementary interfaces. Moreover, since the collection did not change as for P&Z and SL, the combined interface might have had the advantage of being applied to a possibly easier collection – with topics being better distributed or a slightly better working similarity measure so that images of the same topic are found more easily.

Usage of Secondary Focus

For this part, we restrict ourselves to the interaction with the last photo collection where both, P&Z and the lens, could be used and the participants had had plenty of time (approximately 15 to 30 minutes depending on the user) for practice. The question to be answered is, how much the users actually make use of the secondary focus which always contains some relevant images if the image in primary focus has a ground truth annotation. For each image marked by a participant, the location of the image at the time of marking was determined. There are four possible regions: primary focus (only the central image), extended primary focus (region covered by primary lens except primary focus image), secondary focus and the remaining region. Further, there are up to three cases for each region with respect to the (user-annotated or ground truth) topic of the image in primary focus. Table 2 shows the frequencies of the resulting eight

Table 2. **Percentage of marked images (N = 914) categorized by focus region and topic of the image in primary focus at the time of marking.**

focus region	primary	ext. primary	**secondary**	none
same topic	37.75	4.27	**30.74**	4.38
other topic		4.49	13.24	2.08
no focus				3.06
total	37.75	8.75	43.98	9.52

possible cases. (Some combinations are impossible. E.g., the existence of a secondary focus implies some image in primary focus.) The most interesting number is the one referring to images in secondary focus that belong to the same topic as the primary because this is what the secondary focus is supposed to bring up. It comes close to the percentage of the primary focus that – not surprisingly – is the highest. Ignoring the topic, (extended) primary and secondary almost contribute equally and only less than 10% of the marked images were not in focus – i.e. discovered only through P&Z.

Emerging Search Strategies

For this part, again only interaction with the combined interface is analyzed. A small group of participants excessively used P&Z. They increased the initial thumbnail size in order to perceive the depicted contents and chose to display all images as thumbnails. To reduce the overlap of thumbnails, they operated on a deeper zoom level and therefore had to pan a lot. The gaze data shows a tendency for systematic sequential scans which were however difficult due to the scattered and irregular arrangement of the thumbnails. Further, some participants occasionally marked images not in focus because of being attracted by dominant colors (e.g. for the aquarium topic). Another typical strategy was to quickly scan through the collection by moving the primary focus – typically with small thumbnail size and at a zoom level that showed most of the collection but the outer regions. In this case the attention was mostly at the (extended) primary focus region with the gaze scanning in which direction to explore further and little to moderate attention at the secondary focus. Occasionally, participants would freeze the focus or slow down for some time to scan the whole display. In contrast to this rather continuous change of the primary focus, there was a group of participants that browsed the collection mostly by moving (in a single click) the primary focus to some secondary focus region – much like navigating an invisible neighborhood graph. Here, the attention was concentrated onto the secondary focus regions.

User Feedback

Many participants had problems with an overcrowded primary fish-eye in dense regions. This was alleviated by temporarily zooming into the region which lets the images drift further apart. However, there are possibilities that require less interaction such as automatically spreading the thumbnails in focus with force-based layout techniques. Working on deeper zoom levels where only a small part of the collection is visible, the secondary focus was considered mostly useless as it was usually out of view. Current work therefor investigates off-screen visualization techniques to facilitate awareness of and quick navigation to secondary focus regions out of view and better integrate P&Z and SL. The increasing "empty space" at deep zoom levels should be avoided – e.g. by automatically increasing the thumbnail size as soon as all thumbnails can be displayed without overlap. An optional re-arrangement of the images in view into a grid layout may ease sequential scanning as preferred by some users. Another proposal was to visualize which regions have already been explored similar to the (optionally time-restricted) "fog of war" used in strategy computer games. Some participants would welcome advanced filtering options such as a prominent color filter. An undo function or reverse playback of focus movement would be desirable and can easily be implemented by maintaining a list of the last images in primary focus. Finally, some participants remarked that it would be nice to generate the secondary focus for a set of images (belonging to the same topic). In fact, it is even possible to adapt the similarity metric used for the nearest neighbor queries automatically to the task of finding more images of the same to topic as shown in recent experiments [6]. This opens an interesting research direction for future work.

CONCLUSIONS & ACKNOWLEDGMENTS

This paper described the evaluation of a novel interface for browsing multimedia collections in a user study. The results show that the interface is helpful while at the same time being easy and intuitive to use. Based on user feedback, several directions for further development have been identified.

This work was supported by the German National Merit Foundation, the German Research Foundation (project AUCOMA) and the European Commission (contract no. BISON-211898). The authors would further like to thank all participants of the study for their time and valuable feedback.

REFERENCES

1. T. Germer, T. Götzelmann, M. Spindler, and T. Strothotte. SpringLens: Distributed nonlinear magnifications. In *Eurographics - Short Papers*, 2006.

2. M. Lux. Caliph & Emir: MPEG-7 photo annotation and retrieval. In *Proc. of 17th ACM Int. Conference on Multimedia*, 2009.

3. J. Martinez, R. Koenen, and F. Pereira. MPEG-7: The generic multimedia content description standard, part 1. *IEEE MultiMedia*, 9(2):78–87, 2002.

4. S. Stober, C. Hentschel, and A. Nürnberger. Multi-facet exploration of image collections with an adaptive multi-focus zoomable interface. In *Proc. of 2010 IEEE World Congress on Computational Intelligence*, 2010.

5. S. Stober and A. Nürnberger. A multi-focus zoomable interface for multi-facet exploration of music collections. In *Proc. of 7th Int. Symposium on Computer Music Modeling and Retrieval*, 2010.

6. S. Stober and A. Nürnberger. Similarity adaptation in an exploratory retrieval scenario. In *Proc. of 8th Int. Workshop on Adaptive Multimedia Retrieval*, 2010.