3D Interfaces to Improve the Performance of Visual Known-Item Search

Klaus Schoeffmann, Member, IEEE, David Ahlström, Marco A. Hudelist

Abstract—Most interfaces in the field of image and video search use a two-dimensional grid interface, which presents image thumbnails in a left-to-right arrangement that can be browsed from top to bottom. This grid interface, however, has several drawbacks that become particularly apparent when performing interactive search tasks for target items in large collections of images or videos. Therefore, we propose to use 3D interfaces as an alternative to the grid interface for interactive known-item search in visual data as they can partially overcome these drawbacks. In this paper we first summarize our ideas and discuss design aspects of a 3D Ring and a 3D Globe interface. Next, we present results from four different user studies, where we evaluated the performance of these interfaces for known-item search tasks in image collections. Our results from these studies show that the proposed 3D interfaces allow for significantly faster visual target search on desktop computers with mouse interaction as well as on tablet devices. The interfaces also achieve better subjective ratings. However, our evaluation also shows that on smartphones with 3.5-inch screens an improvement over the grid interface in terms of visual search time is only possible in collections with more than 200 images.

Index Terms—Image Search and Retrieval, Video Search and Retrieval, Known Item Search, Image and Video Browsing

I. INTRODUCTION

With photo or video collections, there are situations where users know of a specific item and know that it is contained in the collection, but do not know where to look for it. This is especially true for large personal collections of photos or videos, where users have a clear memory about recorded items and would immediately recognize the desired image or video visually but cannot find it through automatic search (i.e., with retrieval tools). The process of searching for a specific target item in a media collection is also known in the literature as target search or known-item search (KIS) [1]. Image and video retrieval applications [2]–[4] with automatic content analysis and different types for querying could help to find the desired item but (i) users are often not able to appropriately describe their needs through a query and (ii) due to the semantic gap the performance of automatic multimedia retrieval tools is still far from optimal [5], [6]. Therefore, users often have to fall back to interactive visual search. Moreover, even with content-based retrieval tools users typically need to search for the desired image or video in the result list.

Many applications in the field of image retrieval, video retrieval, image browsing and video browsing use a two-dimensional grid interface to display thumbnails of images or key-frames of videos and shots (see Figure 2). For example, most systems used in the interactive KIS task [1] of TRECVID, from 2010 to 2012, employed a grid interface to allow the user to browse results of pre-ranked shots of videos. Similarly, 18 out of the 20 video browsing tools used in the Video Browser Showdown (VBS) [7] competition from 2012 to 2014 – where interactive KIS tasks had to be solved – were based on a two-dimensional interface, 16 of them used a common grid interface as basis. Also many file browsers, image viewers, photo management tools, video editing tools, online image search engines, and video portals use the grid interface to present thumbnails of images and videos.

However, the grid interface has some well-known limitations that become particularly apparent when browsing larger sets of thumbnails. (1) At any time the user can only see a small subset of thumbnails, which makes it hard to see the content structure (i.e., distribution of image characteristics) of the entire collection. (2) With only a small subset of thumbnails visible, comparisons of thumbnails positioned far apart are impossible. (3) Tedious scrolling operations are necessary to navigate through all thumbnails. (4) The interface only allows for a linear search behavior, thus typically, the user has no chance to find an image located in the end of the grid as fast as an image in the beginning. (5) When a user searches through the thumbnails and reaches the end of the list without having found the desired thumbnail, she has to scroll up to the top in order to restart the search from the beginning. Clearly, in this situation the user could also immediately start searching backwards (i.e., in reverse order). However, without a known sorting criterion the user has few, or no, reasons to scroll down and start the search at the end of the list, and most users do not use reverse search [8].

In our previous work [9]–[11], we have proposed and evaluated alternative interfaces that overcome some of the limitations of the widely used grid interface. These interfaces take advantage of a three-dimensional visualization in order to allow for better exploitation of the available screen real estate and, thus, can show more images at a glance in a natural way due to the 3D perspective. Such a 3D interface still has the limitation that only a limited subset of thumbnails is available in one screen, however, this subset is much larger than with a common 2D grid interface. A 3D perspective also enables users to focus their interactive visual search to specific segments of the thumbnail list by viewing some thumbnails at higher level of detail while still keeping unfocused thumbnails in view, though at lower level of detail. This should allow visualizing many images on a single screen, where the user can still focus on a specific area.

One of our proposed interfaces, the Ring interface (cf.
Figure 1a), provides good support for so-called “popout effects”, where distinctive images attract the eye even when they are shown at lower level of detail (e.g., in the background). Another proposed interface, the Globe interface (cf. Figure 1b), allows for horizontal scrolling through thumbnails where images out of focus become smaller and smaller the farther away they are from the center of the screen but remain in view for a longer time than with a grid interface. Both the Ring and the Globe support to immediately restart the visual search from the beginning of the list when the end of the list has been reached without search success. Finally, both interfaces can show more thumbnails at a glance than the grid interface but, of course, have to be carefully configured to not overwhelm the user with too many thumbnails in one screen. In the user studies performed for this work we limited ourselves to collection sizes of a few hundred items, which were still manageable for the tested users.

We tested the 3D Ring and the 3D Globe along with various 2D grid-based interfaces in four user studies [9]–[12] for their convenience and search time behavior. The studies were conducted on three different platforms: a desktop PC, a touch tablet and a smartphone. Figure 1 provides examples of some of the 2D and 3D interfaces we tested. Overall, 100 users participated and we collected performance data from 16,400 search tasks using different data set sizes and search content.

![Image](https://example.com/image.png)  
**Fig. 1.** (a) 3D Ring and (b) 3D Globe. 2D grid-based interfaces (c) and (d).

In this paper, we summarize the results of our studies and discuss the findings from a cross-study perspective. We present additional results on participants’ subjective impressions of the tested interfaces and compare these ratings across all studies. Our results show that 3D interfaces can significantly outperform the grid interface in terms of search time on the desktop PC and tablets, but not for visual target search in small collections with smartphones, where the 3D interfaces perform equally well as the grid. The user ratings indicate that, in general, users perceive the 3D interfaces to be less mentally demanding, less frustrating and more fun to use than the 2D alternative. Our final conclusion from all the studies is that the tested 3D interfaces provide a significant benefit for visual known-item search tasks and, hence, are strong alternatives to the commonly used grid interface. However, in this paper we also outline shortcomings of the proposed interfaces that should be addressed in further studies.

The paper is organized as follows. Section II summarizes related work on 3D interfaces proposed for the field of image and video browsing as well as image and video retrieval. In Section III we summarize the design and implementation of the different 2D and 3D interfaces used in our studies. In Section IV we discuss the results in terms of search time as well as in terms of user rating. Finally, we conclude our work in Section V and outline open questions and further work.

II. RELATED WORK

Several publications can be found in the literature that focus on image/video browsing interfaces with 3D layouts. Most of these proposals are inspired by metaphors from real life. For example, the MediaMetro interface [13] and the mTable interface [14] use a 3D cityscape metaphor to visualize video collections, where single videos are represented by skyscrapers. The RotorBrowser proposed by De Rooij et al. [15] uses a rotor metaphor to visualize similar key-frames of video shots in a 3D arrangement according to several relationships (e.g., visual similarity, semantic similarity, temporal relations in the entire video). The same authors also proposed the CrossBrowser and the ForkBrowser [16], which use a less overwhelming interface for visualizing shot similarity. Through an evaluation in the Interactive Search Task [17] of TRECVID they could show that the ForkBrowser requires significantly less user interaction steps. The VisionGo interface proposed by Theng et al. [18] is based on a horizontal cylinder of key-frames of shots, which can be rotated vertically. Christmann et al. [19] also proposed to use a cylinder metaphor to visualize large image collections. Xu et al. [20] proposed a spiral metaphor to present key-frames of videos and to navigate in videos, where thumbnails are displayed in a vertically aligned 3D helix. Hudelist et al. [21] use the metaphor of a 3D filmstrip to visualize content of videos on tablets. The strip can be zoomed, moved, and dragged in order to navigate through the video. Mueller et al. [22] use a 3D video browser with several 3D layouts for displaying the content of a video, including rotation and zoom interaction. Users can choose whether they want to display video segments in form of a spiral, sphere or cylinder. Slimi et al. [23] use a V-shape 3D arrangement to present key-frames of shots, similarly to the idea of rapid serial visual presentation (RSVP), proposed earlier by Wittenburg et al. [24].

In addition to the above mentioned interfaces that have been proposed to browse videos, several other works focus on the 3D visualization of images. Chun et al. [25] proposed to use a fish-eye visualization to present more images at a glance (similarly, Divakaran et al. [26] did also propose a fish-eye layout earlier). Schaefer [27] proposed to use the globe metaphor for visualizing large image collections with hierarchical refinement, where images are arranged according to their median color. Gomi and Itoh [28] use a 3D visualization to present image retrieval results on a mobile device, where thumbnails are arranged along the Z axis according to their context-dependent priority. A similar idea was proposed by Bruneau et al. [29] for visualization of a clustering graph of images. The photo explorer proposed by Snavely et al. [30] arranges photos in a 3D layout according to geographic coordinates. Finally, Nakazato and Huang [31] utilized a CAVE system to present resulting images of a query according to their similarity in a 3D environment.

It is remarkable that although much effort has been invested in the design and creation of novel 3D visualizations, little effort has been spent on providing empirical evidence that demonstrates their intended advantages over a common grid interface. Likewise, we are also unaware of any other work that evaluates and compares user performance and preferences of
2D to 3D interfaces for interactive visual search with a large enough user study that allows for statistical analysis, which is the focus of this work.

III. TESTED INTERFACES

In order to compare the performance of the 3D interfaces to the performance of the grid interface, we have implemented different versions of them on different platforms. In the following we describe the details of these interfaces as well as considerations that led to specific design decisions.

Color Sorting. Previous work has shown that color-based similarity arrangements can help users in image browsing tasks. For example, Rodden et al. [32] showed that users prefer similarity-arrangements over random arrangements. Later Schaefer [27] also showed that users prefer browsing images with a color layout over browsing images with a name-based arrangement as used in a common file browser. Just recently, we could further demonstrate with a user study that directly compared random arrangement to color-sorted arrangements for a grid interface [9], that visual target search is significantly faster by 20% when the list of thumbnails is sorted by dominant color. Therefore, all interfaces described and evaluated in this paper use a color-sorted arrangement of image thumbnails. It should be noted though that for image and video retrieval tools, it may not be appropriate to present a color-sorted list of query results. In retrieval tools it is rather common to present a ranked list of results. However, we consider our work as a first step towards improving interfaces for visual search tasks. For the retrieval scenario it requires further research to also consider the needs of retrieval tools.

A. The Common 2D Grid Interface

As we were interested in further improving image search performance in color-sorted lists, we used the grid interface (simply referred to as Grid henceforth) as a baseline interface for our studies. Figure 2 shows the different versions of the color-sorted Grid. Part (a) of the figure shows the implementation on a desktop PC with 15.4-inch screen (1440×900 pixels resolution), where the size of the thumbnails was set in accordance with the default thumbnail size in a common file browser. Part (b) and part (c) of the figure depict the implementation on a 9.7-inch Apple iPad screen (1024×768 pixels resolution) respectively on a 3.5-inch Apple iPhone (960×480 pixels resolution). The implementations on the mobile devices were exact visual copies of the default photo browsers on these devices. In all three implementations, thumbnails were arranged by color similarity [33] to allow users to narrow down the visual search for a known-item to a specific area in the list.

B. The 3D Ring Interface

The Ring interface is depicted in Figure 3. The color-sorted list of thumbnails is arranged in circular manner in column-major order using a predefined number of rows and a ring radius. Both parameters are set in accordance with the number of images in the list as well as the available screen space.

For example, the implementation on a desktop PC with a 15.4-inch display typically uses only three or four rows of thumbnails and a rather large radius for a few hundred images, because thumbnails are still recognizable in this setting and using only a few rows produces less occlusion than using more rows. This provides a better view of the entire ring and allows the user to oversee more images. Due to the large screen on a desktop PC the user can still recognize thumbnails in the back of the ring, as shown in Figure 3a for a set of 150 images. However, as the screens on tablets and smartphones are much smaller, we need to use a smaller radius to provide a large enough thumbnail size in the back of the ring (but more rows to show all images). Therefore, our implementations for a 9.7-inch tablet and a 3.5-inch smartphone use three rows only for very few images (e.g., for 100 images as in Figure 3c) but significantly more rows to show a few hundred images. For example, it uses five rows for showing 300-400 images on a tablet, Figure 3b, and six rows for showing the same number of images on a smartphone, as shown in Figure 3d.

The user has free view to thumbnails shown in the back part of the ring. This should allow the user to quickly narrow down the visual search to a specific area in the ring that represents a specific color tone. Furthermore, it should support quick detection of easily recognizable images with highly homogenous color (so-called “pop-out images”). For example, if the user looks for an image with a homogeneous blue background, the image should be easily detectable even if it is in the back part of the ring (compare Figure 3a and b).

The user can directly tap (or click on, when the desktop version is used) a thumbnail to open the full-screen view for the corresponding image. The user can rotate the ring to bring
small thumbnails from the back to the front of the ring. The rotation is controlled by using the mouse-wheel or by swipe gestures on a touch-operated device. Our implementations for tablets and smartphones also support zooming by using a pinch gesture. Moreover, the user can double-tap the back or front of the ring to quickly jump to a fully zoomed view of the tapped area, as shown in Figure 3e and f.

The Ring interface also has an empty column (i.e., a gap) where no images are placed, as visible in Figure 3b. This gap is positioned between the regions of bright and dark images and serves as a visual cue that marks the beginning (and end) of the image collection. Somewhat similar to the position of the scrollbar knob in a 2D interface, this gap – in addition to the color-sorted arrangement – might help the user to build mental anchors for the location of particular images within a frequently used image collection or it might serve as a reference point in situations where the user wants to quickly switch between two or more images in the collection.

C. The 3D Globe Interface

1) Hierarchical Globe (proposed by Schaefer in [27]): Inspired by the work of Schaefer, we have also implemented a 3D Globe interface in order to evaluate it in a user study performed on tablets. As Schaefer describes, the image arrangement in the Globe interface is based on a regular grid structure, where each image is mapped to a cell according to its longitude position (derived from the Hue attribute of the HSV-based median color of the image) and latitude position (derived from the Value attribute of the HSV-based median color of the image). Hence, images are distributed around the longitude according to their median color tone and arranged at the latitude according to their brightness (bright images at north and dark images at south). All images that are mapped to the same cell form a cluster. The cluster center image, which is the image with the minimum areal distance (i.e., distance based on longitude and latitude) to all other images in the cell, is used to represent the whole cluster (see Figure 4a). A hierarchical browsing approach is employed, where the user can click on an image at any level in order to inspect other images in the cluster, if existent. A visual feedback is provided to the user in such a situation by displaying a slightly different view that additionally shows a small representation of the layer above, directly below the visualization of the current layer (Figure 4c). In order to reduce the overlapping, a repetitive algorithm is used, which inspects the 4-neighbourhood of an empty cell that moves across one third of the images from these cells, if at least three of the neighbouring cells are filled (see details in [27]). Moreover, another algorithm is used at lower levels to avoid too deep navigation hierarchies. More specifically, if less than 25 percent of the cells at the current hierarchy level are filled, a circular spreading algorithm is used to distribute images from cells with more than one image over the whole globe. In contrast to the Hierarchical Globe interface originally proposed by Schaefer for a desktop computer, our implementation has been implemented on an Apple iPad and provides pinch gestures for zooming (see Figure 4b). Moreover, our implementation uses a small semi-transparent label for each cluster center image that displays the number of images in the underlying cluster (at the bottom right corner of the thumbnail), as indicated by the inset in Figure 4a.

Fig. 4. Our implementation of Schaefer’s [27] Hierarchical Globe interface on a tablet. Part (a) shows the topmost level of the globe, part (b) the zoomed view of a particular region at the toplevel and part (c) what the interface looks like when the user has navigated into a cluster of the root level.

We have performed a pilot study with the implementation of the Hierarchical Globe interface described above. For that pilot study we invited eight users to perform known-item search tasks in a collection with 350 images. For each task, the user was presented with a desired image (as long as the user wanted to see it) and had to find it as fast as possible afterwards. However, unfortunately it turned out that users had a very hard time in figuring out where an image could be located that is not visible on the root level (i.e., in which cluster). Obviously, the reason for this was that the clustering was not intuitive to the users because we used a small and diverse data set (only 350 images; cf. Figure 4). The Hierarchical Globe interface, however, has been proposed by Schaefer [27] for large-scale image browsing. As a consequence, many images in our comparatively small data size could not be found at all or only after exhaustive search with successively navigating into neighboring clusters. However, at the same time we observed that users were able to quickly find images that were visible on the root level. Therefore, we abandoned the hierarchical clustering feature and implemented two different versions of the Globe interface for non-hierarchical browsing.

2) VaryingGlobe: Another user feedback from our pilot study was that some image positions on the Hierarchical Globe were not very intuitive, especially when considering the brightness of images. Therefore, as a first change we used the Y attribute in the YUV color space (instead of the V attribute in HSV) in order to arrange images vertically, as it showed more intuitive positions according to our observations. Moreover, to avoid hierarchical browsing all images are displayed on the root level. In the first version of this non-hierarchical Globe
TABLE I
OVERVIEW OF ALL USER STUDIES THAT HAVE BEEN PERFORMED TO EVALUATE THE INTERFACES.

<table>
<thead>
<tr>
<th>Study 1 [9]</th>
<th>Interfaces</th>
<th>Platform</th>
<th>Set Sizes</th>
<th>Study Design</th>
<th>Search Time Results</th>
<th>Error Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid</td>
<td>desktop PC</td>
<td>15.4&quot; display touch input</td>
<td>150</td>
<td>12 participants, total 1440 trials each participant performed 60 trials with each interface</td>
<td>Ring significantly faster than Grid</td>
<td>No difference between Ring and Grid</td>
</tr>
<tr>
<td>Ring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2 [10]</td>
<td>Grid</td>
<td>tablet</td>
<td>350</td>
<td>24 participants, total 2160 trials each participant performed 30 trials with each interface</td>
<td>GlobeList significantly faster than Grid (no other differences)</td>
<td>No difference between any interface pairs</td>
</tr>
<tr>
<td>VaryingGlobe</td>
<td>VaryingGlobe</td>
<td>9.7&quot; display touch input</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlobeList</td>
<td>GlobeList</td>
<td>tablet</td>
<td>350</td>
<td>16 participants, total 1280 trials each participant performed 40 trials with each interface</td>
<td>GlobeList and Ring are equally fast</td>
<td>Significantly fewer errors with GlobeList than with Ring</td>
</tr>
<tr>
<td>Ring</td>
<td>Ring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 4 [11]</td>
<td>Grid</td>
<td>smartphone</td>
<td>100</td>
<td>48 participants, total 11520 trials each participant performed 60 trials with each data set in his/her interface group (between-group design)</td>
<td>Across set sizes all interfaces are equally fast</td>
<td>Across set sizes significantly fewer errors with Grid than with Ring or with VaryingGlobe</td>
</tr>
<tr>
<td>GlobeList</td>
<td>GlobeList</td>
<td>3.5&quot; display touch input</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ring</td>
<td>Ring</td>
<td></td>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZoomableGrid</td>
<td>ZoomableGrid</td>
<td>15&quot; display touch input</td>
<td>400</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

implementation (henceforth referred to as VaryingGlobe) the images that are mapped to the same cell due to same median color were displayed at different sizes (see Figure 5a). We used the following rules for determining the thumbnail size: (i) if only one image is mapped to a cell, it is displayed at the normal cell size, (ii) if at least two images are mapped to the same cell, the cell is split into four sub-cells, (iii) if in any sub-cell there are still more than one image (according to the areal position), the sub-cell is further subdivided recursively. As a result, the arrangement of images on the VaryingGlobe consists of image thumbnails of varying size, where images without overlappings are displayed at normal cell size and other images (with overlappings) are displayed at smaller size (cf. Figure 5a). In order to accommodate the different size we provide a zoom-feature that can be invoked through a pinch gesture. However, the usage of different thumbnail sizes is clearly a drawback of this visualization approach because it suggests different (non-existing) priorities of thumbnails. Additionally, it may also produce empty cells if a cell division at any recursion level does not contain exactly four images. As initial experiments with a few users did also not show very satisfying results, we implemented an alternative second version with fixed thumbnail size.

3) GlobeList: In the Globe with fixed thumbnail size (henceforth referred to as GlobeList) we use the same color sorting algorithm but arrange the images according to column-major order, from west to east. In contrast to the Hierarchical Globe, images of the same dominant color tone are not necessarily mapped to the same latitude (cf. Figure 5b). Moreover, bright and dark images are not arranged on the north and south of the globe but shown at the beginning and end of the list (which is initially set to the latitude corresponding to the left part of the screen). As in the Ring interface, to enable the user to better recognize the beginning and end of the list, we indicate it by a column/latitude gap, as visible in Figure 5b.

D. The 2D ZoomableGrid Interface

In order to see if potential differences in the results of our user studies are caused by the intuitive and natural interaction of the 3D interfaces or only due to the fact that they can display more images at once, we designed an additional alternative to the 2D Grid interface for the user study on the smartphone. This interface is basically a Grid interface that first shows all images in one view (i.e., no scrolling is needed; see Figure 6a) and provides a zoom feature to switch to more conveniently sized thumbnails (Figure 6b). The zoom mode is initiated by a double-tap and allows to pan around in the area of the zoomed thumbnails. The gestures for zooming (in and out) and panning are the same as used in common map navigation interfaces on touch-based devices.

Fig. 6. The ZoomableGrid in overview mode (a) and zoom mode (b).

IV. Evaluation

We evaluated the previously described interfaces in a series of user studies [9]–[11] in order to find out whether the designs are beneficial in visual search tasks. We performed four studies, using three different platforms: a desktop PC with mouse input, a tablet and a smartphone with touch input, as outlined in Table I. Here we summarize the results from each study and discuss our findings from a cross-study perspective.

A. Search Time and Errors

1) Study 1: The first user study [9] was performed with 12 users (mean age 30.5, s.d. 3.8) on a desktop PC equipped with a 15.4-inch display. In this study we compared the Grid interface (Section III-A) to the Ring interface (Section III-B). We used a within-group design and tested 120 known-item search tasks for every user (60 with each interface). For each task we showed a thumbnail to the participant and asked him/her to find the thumbnail in a list of 150 thumbnails as fast as possible. The user had infinite time to memorize the thumbnail preview and as soon as he/she felt confident, he/she pressed a button to show the whole list of thumbnails and to start the visual search. The preview of the target image was hidden during the search to avoid simple image matching strategies and to simulate a real known-item search situation,
where the desired image is “in mind” but not visible on the screen. As soon as the user clicked the desired image in the list, the time measurement stopped and the system logged if it was a correct trial. The position of the searched thumbnail varied from trial to trial across the whole list (for more details on the test setup, please see [9]).

The statistical analysis of the data revealed that the Ring interface was significantly faster than the Grid interface by 12.7%, according to a repeated measures analysis of variance (RM-ANOVA) \( F_{1,11} = 6.78, p < 0.05 \). There was no significant difference between the two interfaces regarding the number of erroneous trials where a user found a wrong image for a search task (Ring 7.1%, Grid 7.4%).

2) Study 2: We continued our series of user studies on an Apple iPad (2nd generation) with a 9.7-inch display and 24 participants (mean age 25.8, s.d. 2.9) [10]. We started with a comparison of the Grid interface (Figure 2b) and the two different Globe interfaces (VaryingGlobe and GlobeList) described in Section III-C (Figure 5a and 5b). We again used a within-group design, where each participant had to test every interface. This time we used a collection of 350 images and 30 search tasks for each participant and each interface (cf. Table 1; for more details on the test setup, please see [10]).

The statistical analysis showed that the VaryingGlobe was a little faster than the Grid interface (median search time 12.00 vs. 14.77 seconds), though the difference was not significant. However, the GlobeList was significantly faster than the Grid interface by 23% (median search time 10.16 vs. 14.77 seconds; \( F_{2,46} = 10.09, p < 0.001 \). In total, participants selected the wrong image in 3.4% of the trials. These erroneous trials were equally distributed between the three interfaces (Grid 3.2%, VaryingGlobe 3.7%, GlobeList 3.2%).

3) Study 3: Since we could show that the GlobeList is significantly faster than the Grid interface on tablets, we also wanted to know whether it is faster than the Ring interface too. Therefore, we performed a second user study on the same device (an Apple iPad with a 9.7-inch display) with 16 other users (mean age 27.1, s.d. 3.63) [11]. The study was performed with the same data set (350 images) and a within-group design for the GlobeList and the Ring interface. Each participant performed 40 trials with each interface (for more details on the test setup, please see [11]).

The results showed that users were slightly faster with the GlobeList interface (median search time 7.74 vs. 8.78 seconds). However, an RM-ANOVA did not show any significant main effect for interface \( F_{1,15} = 3.07, p > 0.05 \). In terms of erroneous trials, participants selected the wrong image in 5.5% of the trials when using the Ring and in 2.9% of the trials when using the GlobeList. A Wilcoxon Ranks Test showed that this difference was significant \( Z = -2.2, p < 0.05 \).

4) Study 4: Finally, we performed a user study on an Apple iPod Touch (4th generation), which is equipped with a 3.5-inch display and comparable in terms of capabilities to an Apple iPhone 4/4s [11]. In this study we wanted to test more interfaces and more data set sizes (i.e., longer lists of image thumbnails). The main reason for testing larger set sizes was to find out (i) how the set size correlates with the required search time and (ii) to test if users are still able to manage the interfaces with a few hundred images at one screen (although not the entire collection is visible at once due to occlusions in the 3D arrangement). Therefore, we tested (i) the Grid interface, (ii) the Ring interface, (iii) the GlobeList interface, and (iv) the ZoomableGrid interface. As mentioned in Section III-D, the latter interface was used to evaluate the ‘overview benefit’ that the 3D interfaces provide. This time we used a between-group study design with 48 users (mean age 24.2, s.d. 3.67). Each interface was tested by 12 users. As we tested different set sizes as well (100, 200, 300, and 400 thumbnails), every user had to perform two sessions, because the test would have been too exhaustive otherwise. Therefore, each user tested two set sizes in the first session (with a break in-between) and the remaining two set sizes in the second session. Each user performed 60 trials with each collection (for more details on the test setup, please see [11]).

The analyses showed that the ZoomableGrid interface had a slightly faster search time than the Grid, Ring, and GlobeList interface for the 100 images data set (median search time: 4.6s Grid, 5.6s Ring, 4.8s GlobeList, 4.1s ZoomableGrid) but the difference was not statistically significant. A similar situation was found for the data set with 200 images (median search time: 8.5s Grid, 10.2s Ring, 8.6s GlobeList, 7.7s ZoomableGrid). The situation changed slightly for the data set with 300 images (median search time: 10.9s Grid, 9.6s Ring, 9.2s GlobeList, 9.0s ZoomableGrid) and more clearly for the data set with 400 images (median search time: 13.8s Grid, 10.0s Ring, 9.6s GlobeList, 10.2s ZoomableGrid), with the Grid being slowest and GlobeList being fastest. However, an RM-ANOVA showed no significant effect for the interfaces across all set sizes \( F_{3,44} = 2.71, p > 0.05 \). The number of erroneous trials was similarly distributed between the interfaces for the 100 data set (Grid 1.5%, Ring 2.5%, GlobeList 0.8%, ZoomableGrid 1.7%) and 200 images data set (Grid 1.3%, Ring 3.3%, GlobeList 1.5%, ZoomableGrid 3.3%). For the 300 images data set most errors were registered with the Ring (Grid 1.5%, Ring 8.1%, GlobeList 4.2%, ZoomableGrid 6.9%) and for the 400 images data set most errors were registered with the ZoomableGrid (Grid 1.5%, Ring 5.7%, GlobeList 4.6%, ZoomableGrid 9.2%). The Grid performed quite constant across the four set sizes whereas for the other three interfaces there was a trend with more errors for larger set sizes. A Kruskal-Wallis Test for data across all four set sizes showed a significant difference among the interfaces \( \chi^2 = 11.97, p < 0.01 \) and Bonferroni adjusted Mann-Whitney U Tests showed that significantly fewer errors were made with the Grid than with the Ring \( U = 26.5, p < 0.0083 \) or with the ZoomableGrid \( U = 21.0, p < 0.0083 \). No other differences were detected. From these results we can conclude that the GlobeList performs equal to the Grid on a 3.5” screen in terms of both search time and erroneous trials (across all set sizes).

5) Discussion: Figure 7 shows the median search times for all combinations of interfaces, devices, and set sizes tested in our studies. The results in the figure are grouped by test devices and comparisons should not be made between devices since different input methods (e.g., mouse input in the desktop study and touch input in the other studies) and different screen...
sizes were used (cf. Table I). As visible in Figure 7a and 7b, the 3D interfaces (Ring and in particular GlobeList) allow for significantly faster search for target images in a color-sorted list of thumbnails on desktop PCs and tablets. Especially on tablets the improvement in search time is large: 23% and even larger (47.5%) if we directly compare the results of Study 2 and Study 3 (median search time: 14.77s Grid vs. 7.74s GlobeList on a tablet with 350 images), which both were performed on tablets using the same data set, but with different competing interfaces (cf. Table I). From our findings of Study 3 and Study 4 we can conclude that the Ring and the GlobeList perform equally well in terms of search time when used on a tablet. As apparent in Figure 7c, on smartphones the tested 3D interfaces did only improve search times compared to the color-sorted grid interface for larger data sets (300 and 400 images). Across all set sizes there was no significant improvement for the 3D interfaces (but also no significant degradation). This is obviously caused by the slightly better performance of the Grid with the 100 and 200 data sets. However, when analyzing only the search time with the 400 images, the Grid interface would be significantly slower than all other three interfaces.

On the other hand, our analysis of erroneous trials of Study 4 shows that the Grid was significantly better in terms of wrong image selections than the Ring and ZoomableGrid, but equal to the GlobeList. We speculate that the high number of errors for Ring and ZoomableGrid is due to the very small screen size of the test device we used, which had 3.5-inch screen diameter only. All interfaces but the Grid are designed such that a high number of thumbnails is visible at one glance (compare Figures 5 and 3 and Table II, where the size of thumbnails are listed in relation to the size in the Grid interface). However, with such a very small display the thumbnails become so small that the user cannot recognize them well anymore (for example, with 400 images the thumbnails on the back of the Ring have a size of only 3% of the size in the Grid, as given in Table II). While for many images this is still not a problem and the user is able to quickly find/select it, for other images with many details the user is not able to correctly identify the image in the small thumbnails anymore. The GlobeList – with same performance as the Grid in terms of errors – clearly does not suffer from that problem due to the larger thumbnail size, as also apparent in Table II.

### TABLE II

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Grid</th>
<th>GlobeList</th>
<th>Ring</th>
<th>ZoomGrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 images (3.5&quot;)</td>
<td>100%</td>
<td>82%-27%</td>
<td>46%-11%</td>
<td>100%-20%</td>
</tr>
<tr>
<td>150 images (15.4&quot;)</td>
<td>100%</td>
<td>-</td>
<td>79%- 4%</td>
<td>-</td>
</tr>
<tr>
<td>200 images (3.5&quot;)</td>
<td>100%</td>
<td>42%-12%</td>
<td>20%- 5%</td>
<td>100%-11%</td>
</tr>
<tr>
<td>300 images (3.5&quot;)</td>
<td>100%</td>
<td>30%- 8%</td>
<td>15%- 4%</td>
<td>100%- 7%</td>
</tr>
<tr>
<td>350 images (9.7&quot;)</td>
<td>100%</td>
<td>73%-32%</td>
<td>38%-10%</td>
<td>-</td>
</tr>
<tr>
<td>400 images (3.5&quot;)</td>
<td>100%</td>
<td>23%- 7%</td>
<td>12%- 3%</td>
<td>100%- 6%</td>
</tr>
</tbody>
</table>

From these results we conclude that the 3D Ring allows significantly faster search than the Grid if the screen size is large enough (at least 9.7-inch) but performs equally well on small smartphone screens with 3.5-inch. The same is true for the 3D GlobeList interface, which has the additional benefit of fewer search errors (wrong image selections are comparable to the 2D Grid). From Figure 7 we can conclude that the enabled faster search of the 3D interfaces is due to the higher number of visible thumbnails, because the ZoomableGrid performs equally in this aspect. However, the ZoomableGrid (as well as the Ring) caused significantly more erroneous trials than the Grid on smartphone screens, while the GlobeList did not.

### B. User Rating

In all our studies we also asked participants for their subjective impressions and preferences regarding the tested interfaces. User impressions were collected through questionnaires after all assigned trials were completed with an interface. The questionnaires contained several questions regarding the perceived workload (physical demand, temporal demand, mental demand, required effort, frustration, and perceived own performance) required by the tested interface. In some of the studies participants also rated how much the color-sorting supported them in their tasks (i.e., how comfortable they felt with narrowing down their visual search based on the color) and how fun it was to use the interface. Participants rated each aspect on a 10-point Likert-scale. We based our questions on the workload dimensions suggested in the NASA-TLX (Task Load Index) approach [34]. The NASA-TLX approach is widely used in the field of Human-Computer interaction to assess and compare subjective aspects of interfaces.

In our original work [9] comparing the Grid and Ring for desktop use (Study 1), we did not report on any results about these subjective assessments. Therefore, we describe them in detail here. The mean ratings for each subjective aspect for the Grid and the Ring are shown in Figure 8, where lower ratings are better for all aspects but for ‘perceived performance’, ‘fun’ and ‘support of color sort’. The Ring received a better mean rating on all aspects but on the perceived performance. However, Wilcoxon signed ranks tests only showed statistically significant differences for ‘fun’ \(Z = -2.61, p < 0.01\) and for ‘support of color sorting’ \(Z = -2.09, p < 0.05\). Participants were also asked to rank the two interfaces according to their overall preference; eight users (66%) preferred the Ring and only 4 users (33%) preferred the Grid.

Table III lists the result of the subjective user ratings of the three interfaces examined in Study 2 (Grid, GlobeList, and VaryingGlobe). The first column lists the assessed aspects and the following columns show the results of pairwise interface comparisons on the listed aspects, whereby a \(+\) denotes a
When asked which was their favorite interface, participants were split: 8 voted for the Ring and 8 voted for the GlobeList.

In Study 4, where we compared the performance of the Ring, the GlobeList, the Grid and the ZoomableGrid when used on a smartphone-sized display, we used the same questions. Here we found statistically significant differences between the Grid and the ZoomableGrid for the ‘Effort’ rating, where participants perceived that the Grid required less effort than the ZoomableGrid. The Grid was also regarded to be less frustrating to use than the GlobeList. There were no other pairwise differences between any two interfaces for any other of the used subjective measures. Since we used a between-subject design, there were no favorite rankings in this study.

In summary, considering the user ratings from the three studies where 2D and 3D interfaces were compared (Study 1, Study 2 and Study 4; Study 3 compared two 3D interfaces), we can conclude that, although most of the participants were unfamiliar with 3D user interfaces in general, and none had previously used a 3D interface for image search before, the tested 3D interfaces scored well compared to the already familiar 2D interfaces. Except from the inferior VaryingGlobe in Study 2, all 3D interfaces got equal or better ratings than the tested 2D interfaces on all aspects. However, we are wary of over-interpreting this finding in favor for 3D interfaces in general. Accordingly, we conclude that a 3D-enhancement clearly does not disturb users or negatively effects their overall impressions of the interface.

V. CONCLUSIONS AND FUTURE WORK

In this paper we discussed the findings of four user studies on the performance of 3D interfaces for KIS tasks in image collections. In total 100 persons participated and three different platforms were used (a desktop PC, a tablet, and a smartphone). Our results show that the Ring interface and the GlobeList interface allow for faster visual search times than the 2D Grid on the larger desktop PC and tablet screens. With the small smartphone screen there were no differences between the tested interfaces. In detail, our conclusions regarding the achievable search time are as follows: (i) the Ring is faster than the Grid by 12.7% on a desktop PC with a 15.4-inch display, (ii) the GlobeList is faster than the Grid by 23%-47.5% on a tablet with a 9.7-inch screen, (iii) the Ring and the GlobeList are equally fast on a tablet, and (iv) the Ring, GlobeList, and Grid are equally fast on a 3.5-inch display.

In terms of erroneous trials, the Ring performed equally good as the Grid on a desktop PC but worse on a smartphone screen. The GlobeList, which is equally good as the Grid in this aspect, is the better alternative on small screens.

In terms of user ratings the Ring and GlobeList interfaces perform equally well as the Grid in several aspects but provide significantly more fun for users on the desktop PC and on the tablet. The majority of users (66%) prefer the Ring over the Grid on the desktop PC and the GlobeList over the Grid on the tablet (70.8%). Similar to the findings regarding the search time, there is no difference in terms of ranking for the Ring and the GlobeList on the tablet (50% vs. 50%). Our findings show that 3D interfaces can provide a real benefit for the user on a sufficiently large screen. On desktop PCs and tablets the 3D interface performs equally well as the Grid in several aspects but provides significantly more fun for users on the smartphone.
interfaces can outperform the Grid in both search time and user experience. Further research is required to investigate how to provide the same benefit on smartphone screens. Moreover, our user studies were performed with color-sorted arrangements of image thumbnails. For target images with distinctive colors, the color-sorting enables users to narrow down the visual search to a specific area in the thumbnail arrangement. Further studies are needed to investigate whether a benefit can be achieved with 3D interfaces for ranked lists of images, as used with retrieval tools. We are also interested to test how the different interfaces perform in usage situations other than known-item search situations. For example, in situations where the user does not have a clear target image in mind and instead browses the collection and revisits several candidate images for comparison before a final selection is made. It would also be interesting to study image search performance across a long period of time in order to find out how well the different interfaces foster spatial memory development of image locations in unchanged collections.

VI. ACKNOWLEDGEMENTS

This work was funded by the Austrian Federal Ministry for Transport, Innovation and Technology (bmviit), Austrian Science Fund (FWF): TRP 273-N15, the European Regional for Transport, Innovation and Technology (bmvit), Austrian the different interfaces foster spatial memory development of across a long period of time in order to find out how well the different interfaces foster spatial memory development of image locations in unchanged collections.

REFERENCES

Klaus Schoeffmann is assistant professor at the Institute of Information Technology at Alpen-Adria-Universität Klagenfurt, Austria, where he received his Ph.D. degree in 2009. His current research focuses on visual content analysis and interactive video search. He is the author of numerous peer-reviewed publications on video browsing, video exploration, and video content processing. Klaus Schoeffmann has co-organized international conferences, special sessions and workshops (e.g., MMM 2012, CBMI 2013). He is co-founder of the Video Search Showcase (VSS), formerly known as Video Browser Showdown (VBS). He is member of the IEEE and the ACM and a frequent reviewer for international conferences and journals in the field of Multimedia.

David Ahlström is a Computer Scientist with interests in developing interactive techniques and interfaces. His current research is focused on modeling and empirically measuring human performance with interactive systems with the aim to understand human factors in computer use. He is also interested in around-device interaction techniques for mobile devices. Ahlström is an assistant professor in the Interactive Systems Group at the Department of Informatics-Systems at the Alpen-Adria-Universität Klagenfurt in Austria. He received his M.Sc. degree in Computer and Systems Sciences from Stockholm University and a Ph.D. in computer science from Alpen-Adria-Universität Klagenfurt.

Marco A. Hudelist is a doctoral candidate in computer science at the Institute of Information Technology, Alpen-Adria-Universität Klagenfurt, Austria. He joined the FWF project “Next-Generation Video Browsing” as project assistant in August 2012. In his work he focuses on next generation image and video browsing on mobile devices. He received his B.Sc. degree in computer science in May 2010 and his M.Sc. degree with focus on interactive and multimedia systems in 2012.