A Visual Search User Study on the Influences of Aspect Ratio Distortion of Preview Thumbnails

David Ahlström
Department of Informatics Systems
Alpen-Adria Universität Klagenfurt
Klagenfurt, Austria
Email: david.ahlstroem@aau.at

Klaus Schoeffmann
Institute of Information Technology
Alpen-Adria Universität Klagenfurt
Klagenfurt, Austria
Email: klaus.schoeffmann@aau.at

Abstract—Most image and video retrieval tools used for large-scale media collections present query results as thumbnails arranged in a grid-like display with each thumbnail preserving the aspect ratio of its corresponding source image or video. Often, the outcome of a query is a set of thumbnails with different aspect ratios, thus a varying amount of padding space is used between the thumbnails in the display. This results in a visually erratic display that conflicts with interface design rules and aesthetic principles stipulating alignment and the usage of straight visual lines to guide the human eye while scanning the display. A solution is to create equally sized thumbnails by using cropping algorithms. However, this may remove useful search information. We investigated a simple alternative: to distort thumbnails to the same aspect ratio in order to provide a calm and structured display with straight lines between thumbnails. In a user experiment we evaluated whether and how much such a horizontal distortion can be applied without hampering visual search performance. The results show that distortion does not notably influence error rate and visual search time.

Keywords—image distortion; visual search; image and video retrieval tools; graphical user interfaces

I. INTRODUCTION AND MOTIVATION

Video retrieval applications typically use ‘storyboards’ to show previews of videos in a video archive. A storyboard is a grid-like arrangement of thumbnails of key frames of videos, in which the thumbnails are used as links to the corresponding source videos. Likewise, most interfaces used for browsing personal photo collections or image web-searches display thumbnails in a scrollable list using columns and rows. Allowing fast and accurate access to the video or image source is an important and desirable design goal for most image and video retrieval tools and browsers. For example, finding a particular photo in a large archive can be tiresome and time consuming. Likewise, interactive video retrieval applications, e.g., those used for the Known Item Search (KIS) task of the TRECVID evaluation campaign [1], need to present quick previews in order to enable the user to find an image/key-frame he or she has in mind as fast as possible.

The very diverse set of aspect ratios used nowadays in still camera photography (e.g., 4:3, 3:2, 5:3, 16:9, 16:10) introduces important design decisions to graphical user interface designers who aim for a design that allows the user to quickly visually scan a large set of thumbnails. Presenting each thumbnail with the same aspect ratio as its source guarantees that the user can correctly recognize the content. However, when using a grid as a structuring basis for the thumbnail display the result is a visually erratic display, as visible in the left part of Fig. 1. The varying amount of padding white space between adjacent thumbnails removes much of the grid-structure and provides no straight visual lines that can guide the human eye while scanning the display. As this conflicts with well-established user interface design guidelines and aesthetic principles [2], [3], [4], that encourage the use of alignment and a strive for visual harmony, many designers choose to use image cropping in order to nicely position equally sized thumbnails in a clear grid-arrangement, as shown in the right part of Fig. 1. Beside providing a clear structure, this approach also effectively utilizes precious screen space by minimizing padding white space (particularly important for applications running on mobile devices with small screens [5]). However, cropping may remove important information from an image (especially if the original image has a high aspect ratio), making it hard for the user to recognize the desired thumbnail. Obviously the best solution for equal aspect ratio of all thumbnails would be to extract a square area from each image that contains the most
important information. Unfortunately, for a diverse set of images it is not trivial to find out which region contains most information. For example, while a simple cropping approach that extracts a square from the center of the image produces a quite satisfying thumbnail for Image 1 in Fig. 2, it removes important information of Image 2, as visible in the middle row. A number of researchers have acknowledged the need for more advanced cropping algorithms that detect the most informative region of an image. Approaches include using saliency maps and visual attention models [6], [7], [8], [5] or attention scores based on color and edge features [9]. Others have suggested to use Support Vector Machines trained on thousands of manually cropped images [10]. However, even the best algorithms are unreliable (and often require relatively high computational effort, typically not afforded by search applications such as KIS systems of TRECVID) and will fail for some images. For example, if the algorithm detects the left part of Image 3 in Fig. 2 as the salient area, a very important information is missing in the cropped version (middle row). Similarly, if the right part of Image 4 is detected as the salient area, the cropped version looks quite different.

Given the drawbacks of cropping and of preserving the original aspect ratio for the thumbnails we wanted to investigate the feasibility of a new approach that relies on image distortion to create equally sized thumbnails from sources of varying aspect ratio. By simply stretching or shrinking the height or width of each thumbnail until they all have the desired aspect ratio the thumbnails can be arranged in a harmonic looking and compact space-saving grid-layout, with straight lines of white space between columns and rows, that support fast visual search. Note that for all four example images in Fig. 2 even the extremely distorted version in the bottom row is still quite recognizable and shows a good match with the original image. However, the distortion might impede the user in recognizing the desired thumbnail. As we are unaware of any previous studies investigating implications of image distortion in image search tasks, we designed a user experiment to evaluate whether and how much the distortion introduced by a fixed aspect ratio affects users’ visual search performance. To provide a manageable experimental design and easily interpretable results, we limit this first investigation to horizontal distortion.

II. User Experiment

We designed our experiment according to the visual search paradigm [11] where an experiment participant is prompted with a target item (in our case a thumbnail image) and after inspection has to as quickly as possible determine whether or not the prompted item is present in the following search display. This is also a very similar situation to the Known Item Search (KIS) task of the TRECVID evaluation campaign.

A. Participants, Apparatus and Task

Sixteen volunteers (eight female) aged 26 to 48 years (mean 34.1 years, s.d. 7.3) participated in the experiment. All were frequent computer users and their self-estimated computer usage per week ranged from 14 to 70 hours (mean 45.5 hours, s.d. 18.7).

The experiment was conducted on a Dell Precision M4400 Laptop (running Windows 7) with its 15.4-inch display set at a resolution of 1440×900 pixels. The experiment software was coded in C# .NET 4.0 and its interface is shown in Fig. 3. The software administered 160 visual search trials, each consisting of two steps. In the first step participants inspected a target image that was displayed in the middle of a light grey panel, as shown in Fig. 3a. Participants were told to carefully inspect the cued target image as long as necessary and that after the inspection they were required to search for the same image in a display showing many images without recourse to the cue. After inspection, participants pressed the space bar on the keyboard to proceed to the second step. The second step showed a search display containing 48 images arranged in six columns and eight rows, as shown in Fig. 3b. If the previously inspected image was found in the search display, participants confirmed its presence and the system administered the next trial. If the image was not found, feedback information was displayed to indicate the presence of the target image in the search display.

Figure 3. Experimental interface: (a) the target cue showed in the first task step, (b) the search display containing 48 images showed in the second task step.
presence by pressing the enter key. If the cued image was not present, participants confirmed its absence by pressing the tabulator key. After confirmation (presence or absence), the next trial was loaded and the next target image was cued. Participants were instructed to search the display in the second step as fast as possible. Trial time started when the space bar was pressed and ended when the enter or the tabulator key was pressed (there was no time limit and participants could search the display as long as needed).

B. Experimental Design

The experiment used a within-subjects design with two independent factors: distortion and layout. The distortion factor, the factor of primarily interest, had five levels and controlled the distortion of the images shown in the search display. The images were either undistorted, our baseline, having an aspect ratio of 16:9 (190×90 pixels) or horizontally distorted by a factor of 1.6 to an aspect ratio of 16:10 (144×90 pixels), by a factor of 1.5 to an aspect ratio of 3:2 (135×90 pixels), by a factor of 1.33 to an aspect ratio of 4:3 (120×90 pixels), or by a factor of 1.0 to an aspect ratio of 1:1 (90×90 pixels) – two example images at the five distortion levels are shown in the top part of Fig. 4.

Factor layout had two levels, either packed or padded, and controlled the horizontal width of the search display. Both layouts used a 6×8 matrix to display the images and rows were separated by 6 pixels empty space. The packed layout had a horizontal spacing of 6 pixels between images and horizontally packed the columns around the middle of the search display. Thus, the total width of the search display varied between distortion levels: from 570 pixels (6 image columns×90 pixels + 6 column separators×6 pixels) for aspect ratio 1:1 to 1170 pixels for aspect ratio 16:9. In the padded layout the horizontal width of the search display was kept constant at 1170 pixels by varying the spacing between image columns depending on distortion level: from a 6 pixels padding for the baseline (aspect ratio 16:9) to 126 pixels ((1170 pixels – 6 image columns×90 pixels) / 5 column separators) for aspect ratio 1:1. The bottom part of Fig. 4 shows the visual difference between the packed and padded layout. Together, Figs. 4 a and b visualize that when searching in a packed layout of 90 pixels wide images (a), the eyes have to travel a roughly 50% shorter distance to reach the rightmost image column starting from the leftmost column than when searching in a packed layout with 190 pixels wide images (b). Using more than one single layout in the experiment is important as it helps us to uncover possible effects caused by a particular layout.

The experimental software administered two sessions of 160 trials for each participant, one session for each layout. Eight participants started with the packed layout before proceeding to the padded layout, and eight participants completed the padded layout first. The 160 trials in each session were divided across 10 blocks of 16 trials. Each block contained 10 present trials, two from each distortion level, where the cued target image was present in the search display. In six trials in each block the target image was absent in the search display. The order of distortion level, present and absent trials was randomized in each block. The software ensured that, across the 10 blocks, the 60 absent trials were equally distributed between the five distortion levels. The software also internally divided the search display into four quadrants of 12 images each – upper left, upper right, lower left, and lower right – and ensured that across the 10 blocks each quadrant contained the target image in five of the 20 present trials for each distortion level. Trials terminated with the wrong confirmation key (confirming the presence of an image when the image was absent or visa versa), were logged as errors and were re-queued at a random position among the unfinished trials within the running block. No target position was used more than once for each distortion level in each session.

For each trial, the target image and the images in the search display were randomly drawn from a pool of 1100 images taken from the IACC.TRECVID 2010 repository [1] and the drawn images were distorted according to the current distortion level in case of distortion. An undistorted version of the target image was used as target cue in the first task step. The target image in each trial was removed from the pool after trial completion, all other images used in the trial were put back and served as candidates for the next trial.
The original image pool was restored for each experimental session and participant.

In summary, the number of correctly answered present trials collected in the experiment can be calculated as: 16 participants × 2 layouts (packed, padded) × 5 distortions (1:1, 4:3, 3:2, 16:10, 16:9) × 10 blocks × 2 repetitions = 3200. The number of correctly answered absent trials can be calculated as: 16 participants × 2 layouts (packed, padded) × 5 distortions (1:1, 4:3, 3:2, 16:10, 16:9) × 12 repetitions (distributed across the 10 blocks) = 1920. Each session lasted approximately 20 minutes. Sessions were separated by a 10-minute break. After the last session participants were interviewed about their impressions and experiences.

III. Results

We performed separated analyses for present and absent trials since the two trial types are fundamentally different. Theoretically, if the search image is present in the display, and if the used target positions throughout the experiment are equally distributed across the display (as in our case), the search is, on average, terminated after inspecting half of the images. In absent trials, on the other hand, a complete scan of all images in the search display has to be performed in order to confidentially determine the absence of a particular image. Thus, we expect much faster searches in present than in absent trials.

A total of 5805 trials were performed, 3846 present trials (including 646 errors, 16.8%) and 1959 absent trials (including 39 errors, 2.0%). We identified and discarded 65 present trials (1.7% of all performed present trials) and 29 absent trials (1.5% of all performed absent trials) as outliers, having a trial completion time more than 3 s.d. from the mean for any layout×distortion combination. Outliers were roughly evenly distributed between participants, distortion levels and layouts.

A. Speed-accuracy Tradeoff

Before analyzing error rates and trial completion time in detail, we investigated participants general strategy used when completing the trials by plotting the number of errors made by each participant against his/her mean trial completion time for each of the ten different layout×distortion combinations. The plots revealed that our task, similar to many other cognitive tasks [12], was subjected to a speed-accuracy tradeoff with some participants having a stronger focus on speed, others on accuracy. Across plots, we did not find any marked deviations between any layout×distortion combination for any participant: all participants followed the same, but individual, speed vs. accuracy prioritization throughout the test. We conclude that the adopted strategy was not dependent on neither the layout or distortion factor. Fig. 5 shows each participant’s total number of errors plotted against mean trial completion time across all 10 layout×distortion combinations and makes clear that some participants had a general approach with a stronger focus on speed and others on accuracy. For example, participant 1 and 13 were among the fastest searchers in both present and absent trials. But they did also produce many errors. Thus, indicating an offensive and risky search strategy, buying speed for accuracy. Participant 9, in contrast, exhibits a more defensive strategy with few errors and longer search times. However, most participants seem to have used a balanced strategy. Some were more skilled searchers than others, having both fast trial completion times and few errors (e.g., participant 2 and 11).

B. Error Rate

As expected, we observe much higher error rates in present trials than in absent trials, 18.8% vs. 2.0%, cf. Fig. 6 (note the different scales on the y-axis). Participants’ individual error rates ranged from 5.0% to 28.2% (s.d. 9.6) for present trials and 0.0% to 5.5% (s.d. 4.1) for absent trials. Friedman tests showed that there were no significant differences in error rates between the five distortion levels in neither present nor absent trials (present trials: $\chi^2(4, N=16) = 6.76, p < 0.149$, absent trials: $\chi^2(4, N=16) = 3.34, p < 0.502$). Wilcoxon Signed Ranks Tests showed that there were no differences between error rates for the two layouts in neither present nor absent trials (present trials: $Z = -0.103, p = 0.918$, absent trials: $Z = -0.105, p = 0.916$).

In combination with our previous speed-accuracy analysis, the error analyses show that reducing the width of images by horizontally distorting them did not influence the accuracy with which participants could correctly determine the present trials (1.7% of all performed present trials) and 29 absent trials (including 39 errors, 2.0%). We identified and discarded 65 present trials (including 646 errors, 16.8%) and 29 absent trials (including 646 errors, 16.8%) and 1959 absent trials the presence of a particular image (note the different scales). Error bars show ±2 standard error of the mean.
ence or absence of a target image, and that the frequency with which errors were made was rather influenced by a participant's risk-taking attitude. Participants with an offensive attitude were more likely to quickly terminate a search and thereby take the risk that a present target image was overseen, whereas more careful participants were inclined to spend time to re-scann the search display to confidentially ascertain that the target image was not overseen in case it was displayed.

C. Trial Completion Time

Trial completion time is examined using a 2×5 repeated measures ANOVA with independent factors layout (packed and padded) and distortion (1:1, 4:3, 3:2, 16:10 and 16:9). We analyse present and absent trials separately. The overall mean trial completion time (error and outliers removed) was 2095 milliseconds (s.d. 1368) for present trials and 4561 milliseconds (s.d. 2024) for absent trials.

The mean trial times for all 10 layout×distortion combinations in present and absent trials are shown in Fig. 7. For present trials we see that the packed layout had lower mean trial times than the padded layout at each level of distortion, except from the baseline level (16:9). The ANOVA showed that overall, across distortion levels, the packed layout was significantly faster (9%) than the padded layout (packed: mean 1995 milliseconds, s.d. 1303; padded: mean 2194 milliseconds, s.d. 1423; F1,15 = 13.18, p < 0.01). The ANOVA also showed a significant main effect of distortion (F1,60 = 4.12, p < 0.01), but not a significant layout×distortion interaction effect. Post-hoc pairwise comparisons with Bonferroni adjustment showed that the strongest distortion (1:1) was significantly slower than each of the other distortions (all p’s < 0.05) and that there were no significant differences between any other pairs of distortions. In the right part of Fig. 7 we see that in absent trials the two layouts performed about equally well at all levels of distortions, but at level 1:1 where we see a marked increase in time for the packed layout. The ANOVA showed that this layout×distortion interaction was significant (F4,60 = 3.03, p < 0.05) and that the overall (across distortions) 5% difference in time measures between the two layouts (packed: mean 4435 milliseconds, s.d. 1931; padded: mean 4688 milliseconds, s.d. 2105) was not significant. The ANOVA also showed a significant main effect of distortion (F2,8,42.8 = 4.71, p < 0.01, Greenhouse-Geisser corrected) and post-hoc pairwise comparisons with Bonferroni adjustment revealed significant differences between the strongest distortion level (1:1) and both the baseline (16:9) and distortion level 16:10 (both p’s < 0.05).

These analyses show that the time it took to visually search for an image when the image was present in the search display was not hampered when the images were distorted, as long as the distortion did not exceed a certain threshold. With our setup, the threshold appears somewhere between distortion 4:3 and 1:1.

The significant effect of the layout factor in present trials shows that the larger search area in the padded displays and their padding white space slowed down the search. This confirms previous studies reporting that less dense search displays can be disadvantageous, e.g., [13], [14]. The absence of a significant layout×distortion interaction for present trials shows that the disadvantage of a padded layout was equally pronounced at all distortion levels.

The analyses also show that if the distortion was no stronger than 4:3 it did not markedly impede participants in situations where they had to ascertain that a particular image was not visible in the search display. However, the significant layout×distortion interaction for absent trials shows that the exceptionally bad performance with the most distorted images is caused by the padded layout alone, presumably as participants had to struggle with both a large display and strong distortion.

D. Subjective Impressions

When interviewed after the experiment, only ten participants (62.5%) claimed to have noticed “some changes” in the images or that some images were “squashed” or distorted. Only four of these ten participants believed there were more than two levels of distortion and only two participants said that they sometimes found the distortion disturbing. These subjective impressions are quite remarkable considering the fact that in only 25% of the trials the images in the search display had the same aspect ratio as the previously shown target cue. In other words, even the strongest distortion level (1:1), where the width of images was crushed to less than the half, was not apparent to more than 37% of the participants.

Nine participants stated a clear preference for the packed layout, two participants favored the padded layout, and five had no preference.

IV. CONCLUSION AND FUTURE WORK

Since our experiment participants were able to find the target images in the displays with distorted images as fast as they could in the displays with undistorted images, we conclude that the distortion did not make it harder for them to recognize the cued images. This demonstrates that image
distortion does not necessarily hamper user performance. In our case, images with an original aspect ratio of 16:9 could be distorted down to an aspect ratio of 4:3 – a reduction in image width by 36.8% – without time sufferings or higher error rates. Our results also demonstrate that the idea to use distortion in video or image retrieval tools (or any other interface that needs to display a large number of images on a limited screen space) is a feasible and promising approach to create equally sized thumbnails in order to achieve a visually harmonic result display with straight columns and rows of thumbnails – which supports visual search. The main advantage of using distortion is that it eliminates the risk to lose vital visual information, which is likely to happen when an automatic cropping algorithm is used. Furthermore, not only conventional two-dimensional image browsing interfaces require thumbnails to be at a different aspect ratio than the original image, also more novel interfaces using 3D graphics, such as cylindrical displays [15], [16], or spherical ones [17], rely on equal aspect ratios in order to arrange their thumbnails.

Our experiment was limited in that it only used images with landscape orientation and horizontal distortion. Thus, to provide a broader understanding of the implications of distortion on image recognition and visual search performance, further user experiments are needed which include portrait-oriented images, vertical distortion and distortion caused by perspective changes. We are also interested in investigating situations where images are distorted through enlargement.

ACKNOWLEDGMENT

This work was supported by Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant KWF - 20214 22573 33955.

REFERENCES


