Visual Informatics 1 (2017) 16–24

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Image grid display: A study on automatic scrolling presentation

Marco Porta *, Stefania Ricotti

Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

ARTICLE INFO

Article history:
Available online 18 January 2017

Keywords:
Image grid display
Automatic scrolling
Image presentation modes
Image collections
Rapid serial visual presentation

ABSTRACT

In this paper we describe a study on image grid display with automatic vertical scrolling. While scroll operations are normally carried out manually by the user, in the context of RSVP (Rapid Serial Visual Presentation) techniques this work considers a presentation mode in which the image grid is automatically scrolled. Through experiments carried out with 50 testers, we have investigated user performance while looking for specific target subjects within large collections of images. Different numbers of columns and scrolling speeds have been considered. The search task implied both clicking on the identified target pictures and simply vocally stating their visual recognition. To this purpose, and to identify possible specific gaze behaviours, eye tracking technology has been exploited. The obtained results show that number of columns and scroll speed do affect search performance. Moreover, the user's gaze tends to focus on different screen areas depending on the values of these two parameters. Although it is not possible to definitely find an optimal columns–speed combination that is valid in all cases, the particular context of use can suggest feasible solutions according to one's needs. To the best of our knowledge, image grid display with automatic scrolling has never been studied to date.

1. Introduction

Grid (or tabular) display is certainly the most common way of presenting groups of images on a screen. For instance, so-called "thumbnail" pictures arranged in rows and columns can be found in online previews (e.g. result pages of 'Google image search' or photos in Instagram profiles), folder content display (e.g. MS Windows medium, large, and extra-large icons), or graphic tools (e.g. in the 'Styles' palette of Adobe Photoshop). The main advantage of the grid is evident: since images are normally rectangular, it allows an efficient exploitation of the available space. It is also intuitive, as the table layout is used to display many kinds of data, outside the context of pictures as well.

Image grids come in a variety of configurations, in which image size, number of columns and number of rows are the most common variables. Of course, the smaller the pictures the higher their display density on the screen. However, small graphic representations may be difficult to identify, especially when the recognition of specific images or subjects is required. A trade-off is therefore necessary between the number of presented images and their size, which usually depend on factors such as presentation goals, kind of subjects, and needed search accuracy.

Unless the number of pictures is very low, vertical scroll operations are almost always necessary. Alternatively, different sets of images can be displayed in separate "screens", as sometimes happens in websites. However, especially online, loading new screens may result in a slower image browsing experience, which may not be always acceptable. A combination of the two solutions—different screens containing scrollable grids—is an adequate choice in many cases.

In this paper, we consider an unusual approach to image grid display with vertical scrolling. While scroll operations are generally performed manually by the user, here we focus on a presentation mode in which the grid is *automatically* scrolled up with a constant speed. In other words, the user does not need to employ the mouse or the keyboard to move the displayed set of images, which is automatically shifted upwards. The rationale for this is twofold.

Firstly, an automatic scroll can be seen as a simulation of a manual scroll with constant scroll rate. Thus, indirectly, it is possible to study user behaviour while performing one of the most common—albeit typically "unconscious"—operations in grid display with many images. For experimental purposes, using the same speed with all testers can provide more reliable results (at least for the goals of our investigation).

The second and most important reason for which we decided to study scrolling image grids relates to *Rapid Serial Visual Presentation* (RSVP) modes (Spence, 2002). Putting it simple, RSVP

* Corresponding author.
E-mail address: marco.porta@unipv.it (M. Porta).
Peer review under responsibility of Zhejiang University and Zhejiang University Press.

http://dx.doi.org/10.1016/j.visinf.2017.01.003

means displaying many images in rapid succession on the screen. The purpose of this visualization strategy is to present numerous pictures in a short time, so that users can rapidly find what they are looking for. While in *indexed* image databases pictures have textual information associated with them, and in *Content-Based Information Retrieval* (CBIR) systems pictures are identified based on intrinsic characteristics such as colour, shape and texture, in many cases the manual browsing of very large collections of images is necessary. In fact, there are several situations in which neither indexed nor CBIR systems are available, or we need to select only some pictures because they are characterized by specific features or simply because we "like" them (e.g. for decorative purposes). Browsing big collections of images with traditional display modes—e.g. the static grid—may result in a very boring task, and RSVP can be the right solution.

Several variants of the basic RSVP method (in which single pictures, one at a time, are rapidly displayed on the screen) have been proposed (Cooper et al., 2006; Porta, 2006, 2009). For example, in the *mixed* presentation mode images are displayed in groups of four or more; in the *diagonal* mode, images move from a corner of the screen to the opposite one; in the *collage* mode, pictures appear in random positions of the screen, like if thrown on a table; in the *volcano* mode, images emerge from the Centre of the screen and move radially towards the edges; in the *fountain* mode, pictures are randomly "spurted" upwards and then fall back.

Although some general guidelines have been drawn from experimental evidences (Witkowski and Spence, 2012; Spence and Witkowski, 2013), each RSVP solution is characterized by its own features and needs customized investigations for its potentials and drawbacks to be really understood.

To the best of our knowledge, image grid display with automatic scrolling has never been studied to date. In this paper, we therefore focus on this presentation mode and investigate its possible use as an RSVP method. In particular, we consider two main variables that affect its performance, namely *number of columns* and *scroll speed*. While also image size may be an important factor, in our experiments we opted for a constant value, so as to limit the complexity of the analysis. Observing the heights of pictures displayed, for example, in the result pages of the main search engines (which range from about 140 pixels of Bing to about 170 of Yahoo! and 180 of Google), in MS Windows large icon display (90), or in YouTube home page video thumbnails (110, 170 including the description), we can notice a relatively high variability. Even if our choice – 150 × 150 pixels—cannot account for all possible dimensions, we think it is a good compromise between widespread sizes and acceptable accuracy in the recognition of image subjects.

For our analysis, we also exploited *eye tracking* technology (Duchowski, 2007). An eye tracker is a device capable of recording one's gaze direction while looking at a screen, thus allowing precise understanding of what the user is watching during experimental tests. Eye movements occur as very fast *saccades* followed by *fixation* periods of about 100–600 ms, during which the eye is almost still. Eye tracking data have provided us with interesting insights into gaze behaviours and their relationships with the considered variables.

In summary, this paper tries to answer two main research questions regarding image grid display with automatic scrolling:

- Q1 Is there any best combination of number of columns and scroll speed that guarantees a "good" performance when the purpose is to find specific subjects within an image set?
- Q2 Is image search characterized by any particular gaze behaviours? And, if so, do these behaviours depend on number of columns and scroll speed?

The article is structured as follows. In Section 2 we describe some previous works directly or indirectly connected with grid layout image display and RSVP presentation modes. In Section 3 we then illustrate our study, in terms of participants and task description. Subsequently, in Section 4, the obtained results are presented in relation to user performance, gaze behaviour and gaze scan path length. Lastly, in Section 5, we discuss the achieved outcomes and draw some conclusions, also providing hints for future research.

2. Related work

Although there seems not to be any study specifically devoted to scrolling image grid display—let alone investigations focused on automated scroll—there are works focused on variants of the grid display, rapid serial visual presentations, and the use of eye tracking as an evaluation tool or an active control mechanism for information visualization: here we provide some representative examples of them.

PhotoFinder (Shneiderman and Kang, 2000) and PhotoMesa (Bederson, 2001) are two cases of early and well-known visualization techniques based on the grid. The first is a photo annotation tool that allows the user to add captions and edit images, while the second is an application in which multiple directories of images can be viewed within a zoomable environment. Solutions have also been proposed (Igarashi and Hinckley, 2000) which dynamically control grid zooming depending on the scrollbar speed (the faster the scrollbar, the less the zoom applied): since the grid display usually requires scrolling, small movements of the scrollbar may produce big shifts of the grid if it contains many images.

Often, pictures are also clustered according to some criteria in more or less standard grid arrangements, so that they can be hierarchically browsed in non-linear manners or arranged according to their mutual similarity. For instance, Liu et al. (2004) analyse user needs for web image search and propose a similarity-based organization to present search results. Ren et al. (2009) describe an interactive interface in which images are clustered through an unsupervised graph-based clustering algorithm. Strong et al. (2010) present an approach supporting dynamic zooming in which visually similar images are displayed in the same locations (either scattered or aligned to a grid, depending on the selected display mode). Kleiman et al. (2015) describe a system in which images are dynamically arranged (on-the-fly, depending on users' navigation tendencies and interests) close to their nearest neighbours in a high-dimensional feature space.

A variant of the pure grid arrangement is presented in a work by Schaefer (2010), where a similarity-based picture organization approach is used to display images onto a sphere, with which the user can interact. Likewise, a 3D interface is used by Schoeffmann and Ahlström (2012) to display image thumbnails on a cylinder, based on visual similarity.

As regards RSVP approaches (some of which connected with eye tracking), Fan et al. (2003) developed a prototype solution for browsing large amounts of images which exploited a gaze-based attention model. Oyekoya and Stentiford (2006) studied a gaze driven image retrieval system and tested it with 13 users who had to find target images within 4 × 4 screens, with the screen automatically changing when the duration of all fixations on a picture reached a threshold. Results indicate a slower mouse response compared to the eye tracking approach. Corsato et al. (2008) compared four RSVP techniques (Floating, Collage, Volcano, and Shot) in the specific task of finding the highest number of pictures matching a textual description, similarly to our experiments. The 30 testers involved were studied through an eye tracking system with the main aim to validate the viability of a gaze controlled image selection method. Besides ordinary pictures, other studies

have also considered the use of RSVP as a means for browsing or searching large amounts of data like Beck et al. (2012), who used RSVP to visualize graph data diagrams evolving over time. In a recent work, van der Corput and van Wijk (2016) proposed a system that provides two different visualization modes, namely "pivot" and RSVP, the latter being optimized for detailed inspection and fine-grained categorization of images arranged in a grid.

3. Our study

We conducted 125 tests in which participants had to find a specific target subject within a set of 250 random images. The controlled variables were the *number of columns* and *grid speed*. Five target subjects were employed, namely trains, airplanes, cars, motorcycles, and ships. Besides using eye tracking technology to study user behaviours, participants were also video and audio recorded during the tests, and were instructed to both click on target images with the mouse and tell out loud that they saw them, as soon as they could: this allowed us to correlate gaze positions with the real will to select an image, even when the grid moved too fast for actual mouse selection.

3.1. Participants

Fifty testers (28 females and 22 males) participated in the study: 44 of them were aged between 20 and 25, five between 26 and 30, and one between 40 and 45. All had normal or corrected to normal vision (11 were wearing glasses).

3.2. Task description and procedure

We created five sets of target images, each one formed of ten pictures representing a means of transport (as said, trains, airplanes, cars, motorcycles, and ships).

Each tester was presented with five different moving grids of images (created using Adobe Flash), each with a different combination of number of columns and scroll speed, as well as containing a different target subject (10 target images in total). For each presentation, participants had to find target pictures pertaining to one of the above mentioned five target sets: for instance, the request to the tester could be "Find images depicting a car".

Each presentation was generated by a Flash script in the following way:

- Images were randomly selected from a set of 3219 general photos extracted from various image collections, both public (e.g. Microsoft Office cliparts) and not. None of these photographs contained trains, airplanes, cars, motorcycles, or ships;
- Three parameters, namely SPEED, COLUMNS and SET, were associated with each combination of number of columns and speed to specify, respectively, number of columns, speed, and target image set;
- Each presentation included 250 pictures (240 + 10 target images, randomly arranged). The grid initially entered the screen from the bottom and finally disappeared through the upper edge. However, to avoid that some target subjects could appear in the first or last part of the presentation (when less images were displayed on the screen, because the grid was starting or ending its motion), n pictures were added to the 250 images both at the beginning and at the end of the grid, where $n = \text{COLUMNS} \cdot 6$ (and 6 was the number of rows filling a screen). These images, randomly chosen from the set of 3219 photos, were different from the previous 240, and did not contain any of the five target subjects. This way, the actual 250 pictures considered in the tests always completely filled the screen.

With regard to presentation speed, the grid moved upwards by SPEED pixels at each "execution" of the Flash frame. Since we used a frame rate of 25 fps, the actual presentation speed was $25 \cdot \text{SPEED}$ pixels/second.

The Flash program automatically counted the number of target images correctly selected with a mouse click, the number of misplaced clicks (i.e. mouse clicks on images which were not target images), and the list of target images correctly selected.

The size of each image was 150 × 150 pixels, and both the vertical and the horizontal space between adjacent images was 24 pixels. The number of columns (COLUMNS) could vary from 3 to 7. Seven was the maximum number of columns compatible with the eye tracker screen width (1280 pixels), and the fixed width of images and gaps between them. Three was chosen as the minimum number of columns since grid image arrangements that can be found on the Web usually have at least three columns.

Fig. 1 shows the five presentation modes, ranging from three to seven columns.

The scrolling speed (SPEED) could assume five values, namely 5, 10, 15, 20, or 25—which correspond, respectively, to 125, 250, 375, 500, and 625 pixels/second. We expected 5 and 25 to be "extreme" values, respectively very slow (125 px/s) and very fast (625 px/s), to be used as a basis for comparison with the other three "midway" solutions.

Testers were instructed to both click on target images and say out loud the word "seen" as soon as they saw one of them, so that, even when the scrolling speed was too high to precisely click on a target image, we could know (in the subsequent analysis of eye tracking videos) whether the image was actually fixated or not. We considered as "valid" all fixations that were detected inside a circle with a radius of 140 pixels centred on the image—i.e., in practice, within the circle circumscribing the square obtained by adding a 24-pixel frame around the target image. This extra space (that was the gap between adjacent pictures) was considered because, especially at high speeds, an error was introduced by the gaze recording software, which was unable to record all the necessary video frames. Some preliminary experiments allowed us to determine that a 140-pixel radius circle (which we will call *target circle* in the following) is enough to compensate for this error.

All possible combinations of number of columns (3, 4, 5, 6, 7), speed (5, 10, 15, 20, 25), and target subjects (trains, airplanes, cars, motorcycles, ships) were covered, using a 25 × 25 square design. Each of the 25 COLUMNS–SPEED combinations was tested 10 times, with only one exception for "COLUMNS 7" and "SPEED 15", because, due to a technical problem, the test was not correctly recorded. Therefore, we collected a total of 249 recordings from 50 testers.

As an eye tracker, we used a *Tobii 1750*, which incorporates its gaze tracking sensors and near-infrared light emitters into a 1280 × 1024 pixel (17") screen. The *Tobii Studio* software was employed to implement the experiments and record gaze data.

4. Results

The following statistical analysis was carried out using both parametric and non-parametric statistics (5% significance level), depending on the results of the Kolmogorov–Smirnov test conducted on the different sets of data to determine whether the distributions were normal or not.

In the analysis, we considered both *clicked* target images, i.e. the pictures that were actually clicked with the mouse, and *seen* target images, i.e. all the target pictures (clicked or not) testers claimed to have detected by saying "seen". These detections were subsequently confirmed or denied by checking the gaze replays of all 50 testers, to verify that when they said "seen" their fixations were actually located in the target circles circumscribing target images.

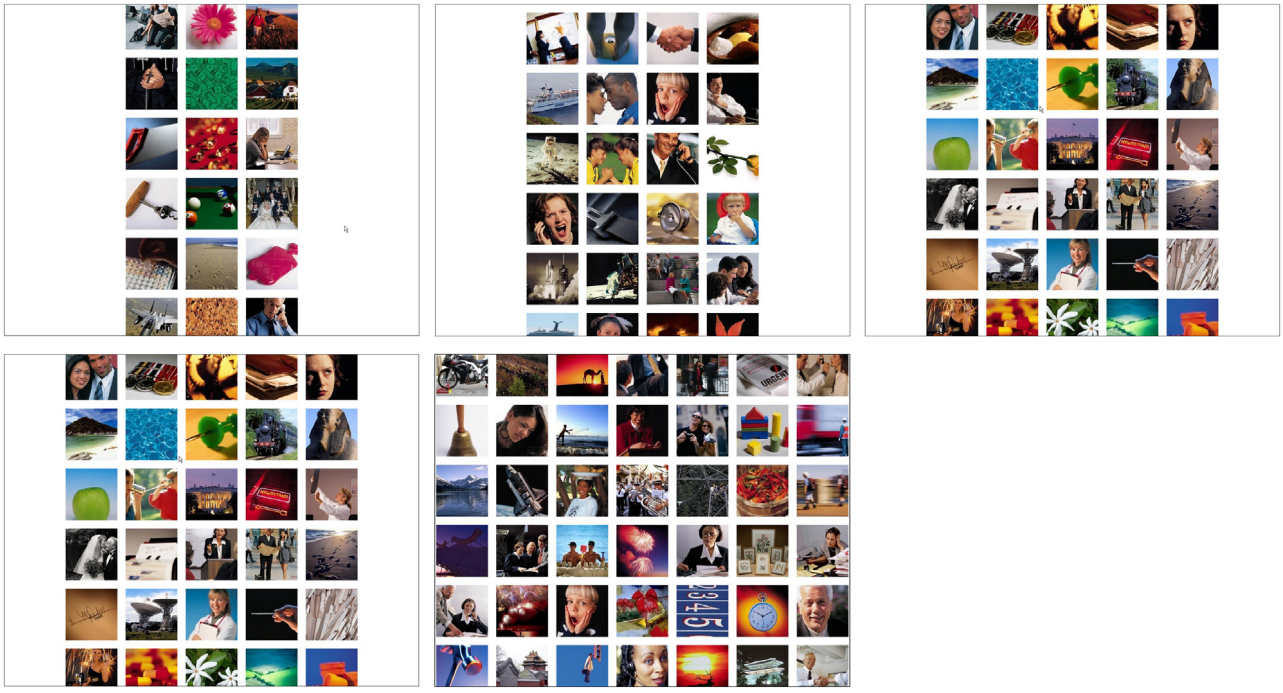

Fig. 1. Image grid with three, four, five, six, and seven columns.

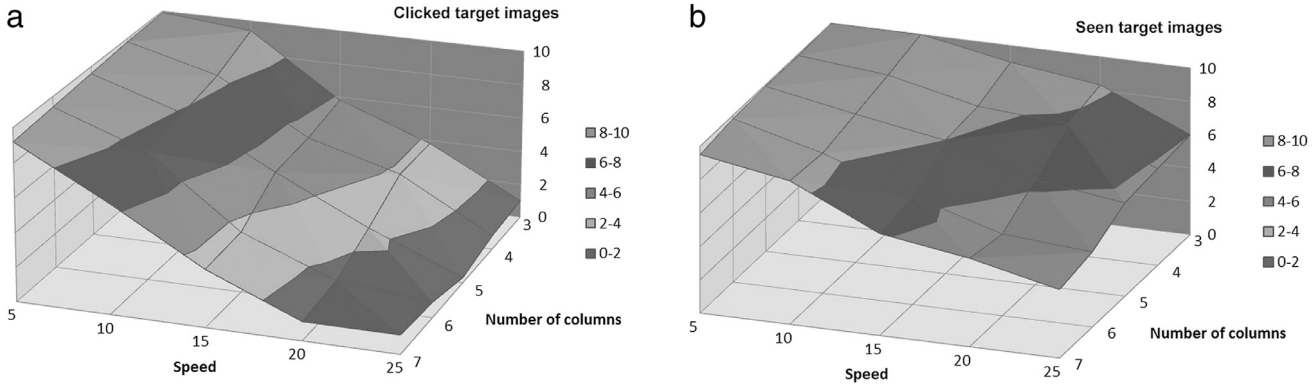

Fig. 2. Average number of clicked (a) and seen (b) target images depending on number of columns and scrolling speed.

Of course, our assumption is that the specific target subject (trains, airplanes, etc.) did not affect test results. To this purpose, we calculated the means of the numbers of correctly seen images for each target subject and for each one of the 25 COLUMNS–SPEED combinations. The differences of these means for the different targets were very small: calculating their standard deviations, still for each COLUMNS–SPEED combination, in 11 cases they were less than 0.5, in five cases between 0.5 and 1, and in the other nine cases between 1 and 1.5. Moreover, none of the target subjects was always seen more or less than the others over the different combinations.

4.1. User performance

Fig. 2 shows users' performance with the different COLUMNS–SPEED combinations for correctly clicked (a) and seen (b) target images. Fig. 3 shows the same information in a single histogram, also indicating the number of images wrongly clicked. Fig. 4 shows the number of clicked and seen target images along with their "discrepancy", based on number of columns (a) and speed (b).

As can be seen from Fig. 2, Fig. 3 and Fig. 4, and could be reasonably expected, number of columns and scrolling speed do affect both the number of seen and clicked target images. However, while both these values decrease similarly when the number of columns increases (Fig. 4(a)), the number of clicked images decreases much more rapidly with the increase of scrolling speed (Fig. 4(b)): although the user may be able to notice the target image, at high speeds it becomes very difficult to click on it—the "reaction time" cannot cope with the fast scrolling grid.

Regarding the error rate, as can be seen from Fig. 3, the mean number of wrongly clicked images is always very low (all target images were very well recognizable).

Considering the number of columns only, no significant correlation was found between number of columns and wrongly clicked target images: Pearson's $r = .43$, $p = .47$. Also for scrolling speed, no correlation was found, although the result in this case was close to significance (Pearson's $r = .859$, $p = .062$).

From a statistical point of view, we inferred that there is a significant correlation between number of columns and clicked target images: Pearson's $r = -.99$, $p < .001$. Likewise, there is a significant correlation between grid speed and clicked target

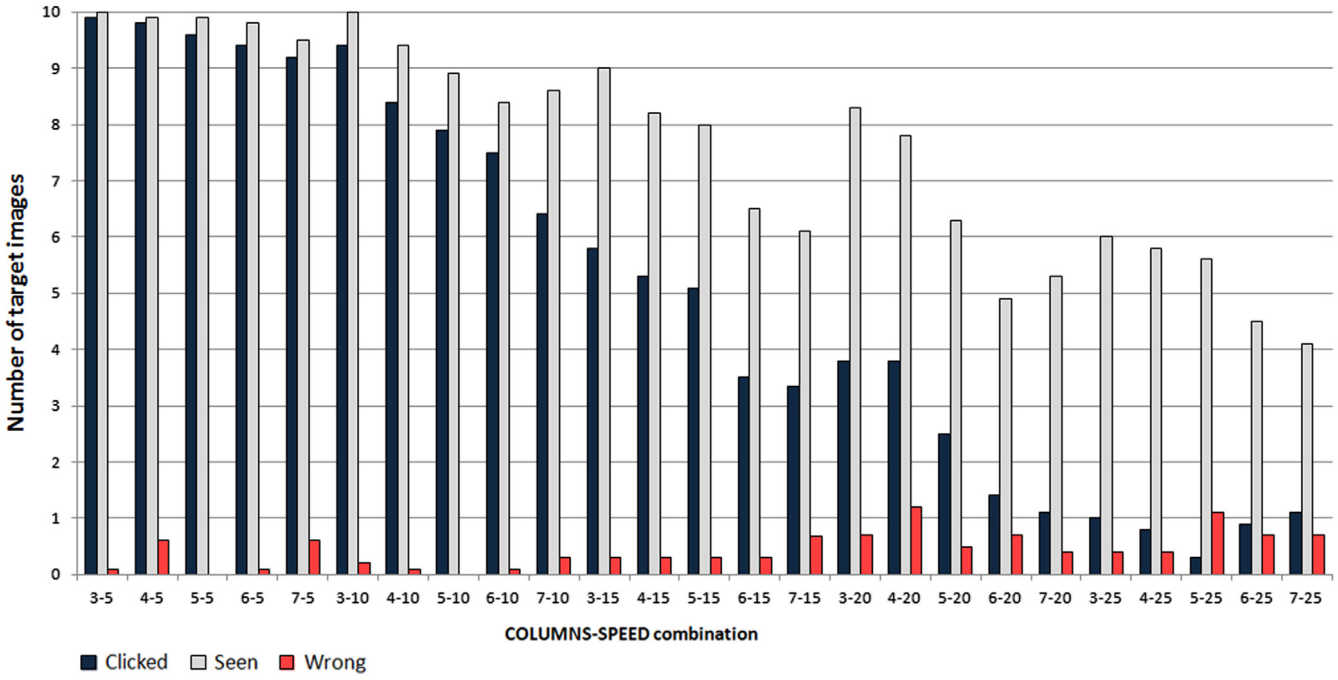

Fig. 3. Average number of clicked, seen, and wrongly clicked target images for each combination of number of columns and scrolling speed.

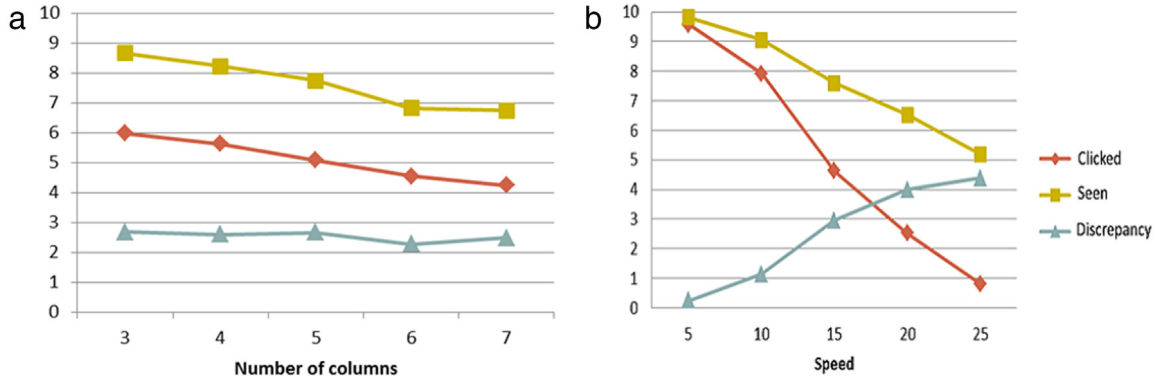

Fig. 4. Average number of clicked and seen target images and their difference, based on number of columns (a) and speed (b).

images: Pearson's $r = -.99$, $p < .01$. Similarly, there is a significant correlation between number of columns and seen target images: Pearson's $r = -.98$, $p < .01$. There is also a significant correlation between grid speed and seen target images: Pearson's $r = -.99$, $p < .001$.

4.2. Gaze behaviour

To study the distribution of testers' fixations on the scrolling grid, we vertically divided the screen into three equal areas, which we will simply call *Top*, *Centre*, and *Bottom*. For each recording, we then determined which of the three areas received most fixations.

We found that almost all participants tended to focus their gaze on the central and lower parts of the screen, while in very few cases (just nine out of 249 recordings, 3.6%) testers tended to concentrate their fixations in the upper part of the screen. In 96% of cases, the most watched area contained at least 50% of fixations, and in 61% of cases more than 60%. Only in 30 cases out of 249 (12%) the most watched area included fewer fixations (21 between 45% and 50%, 8 between 40% and 45%, and one between 35% and 40%). Fig. 5 shows the distribution of fixations among the three regions, considering, respectively, only the variable 'Number of columns' (a) and only the variable 'Speed' (b).

Fig. 6 shows three examples of *gazeplots* (graphical representations of sequences of fixations, depicted as circles with areas proportional to their durations) with fixations mostly concentrated, respectively, in the Bottom (6(a)), Centre (6(b)), and Top (6(c)) areas.

As can be seen from Fig. 5, with few columns (less than 6) most testers tended to look at the bottom of the screen, while fixations tended to be concentrated in the central area when the number of columns increased. Analogously, as regards scroll rate, when the speed was not high (less than 15) testers tended to look at the bottom of the screen, while with increasing speeds fixations tended to be concentrated in the central area. Indeed, there was a significant correlation between the number of columns and the number of testers who focused their gaze on the central (Spearman's $rho = .47$, $p < .05$) and lower (Spearman's $rho = -.52$, $p < .01$) parts of the screen. There was also a significant correlation between the presentation speed and the number of testers who focused their gaze on the central (Spearman's $rho = .60$, $p < .01$) and lower (Spearman's $rho = -.60$, $p < .01$) screen areas.

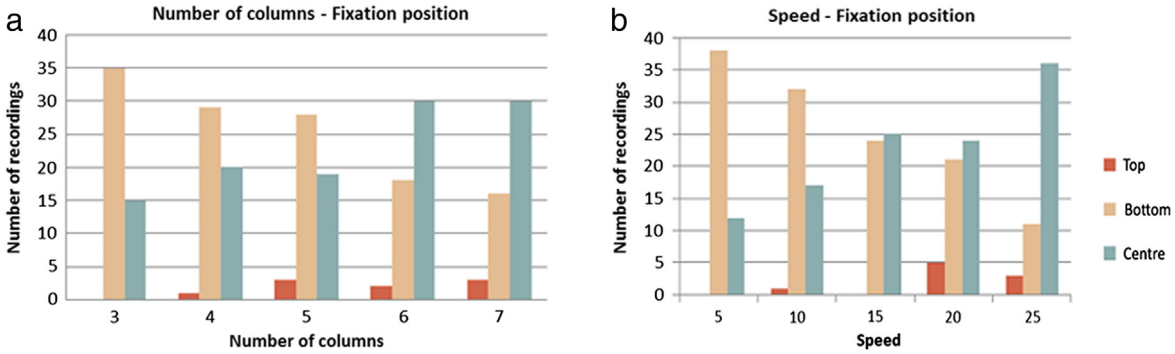

Fig. 5. Distribution of fixations among the Top, Bottom, and Centre areas according to number of columns (a) and speed (b). For each value on the X axis, the histograms indicate, for each area, the number of recordings (i.e. presentations) for which most fixations were detected in that area.

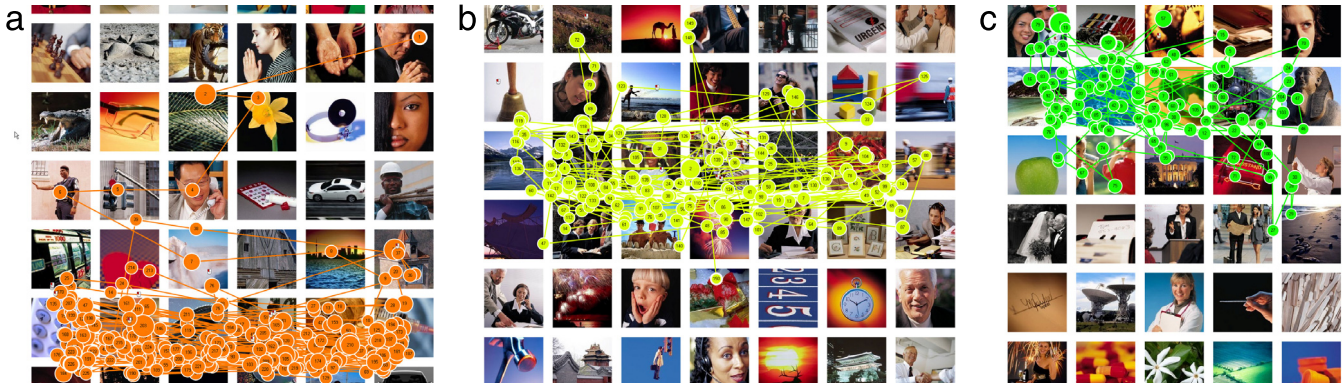

Fig. 6. Examples of *gazeplots* in which fixations are mostly concentrated in the Bottom (a), Centre (b), and Top (c) areas of the screen.

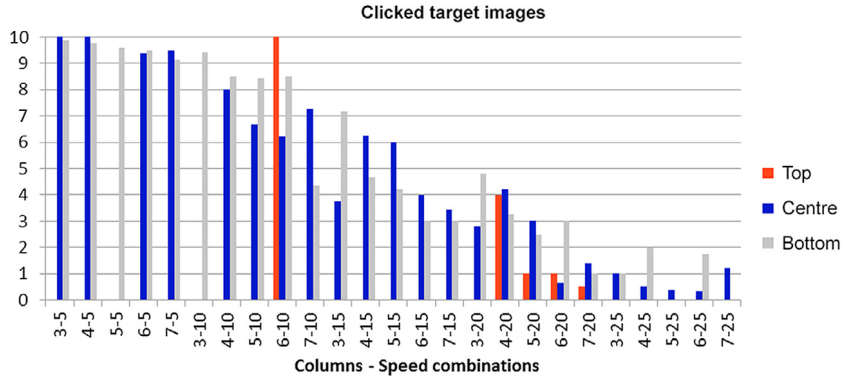

Fig. 7. Average number of target images clicked by testers who mostly focused their gaze on the Top, Centre, and Bottom areas.

Considering solely the Centre and Bottom areas, which received almost all fixations, we also investigated whether the position of fixations has a relation with the performance of testers. Among the 25 possible combinations of number of columns and presentation speed, we considered only the 21 cases in which there were both testers who watched the Centre region and testers who watched the Bottom region. From the histogram in Fig. 7, we can see that the average number of target images clicked by testers who focused their gaze mostly on the central part of the screen is higher than the average number of target images clicked by testers who focused their gaze mostly on the central area for 11 combinations out of 21, while the opposite occurs nine times (and in one case the two numbers are the same). Analogously, from the histogram in Fig. 8, we can notice that the average number of target images seen by testers who focused their gaze mostly on the lower part of the screen is higher than the average number of target images seen by testers who focused their gaze mostly on the central area for 10 combinations out of 21, while the opposite occurs eight times (and in three cases the two numbers are exactly the same).

However, using independent samples *t-tests*, we inferred that the above differences are not statistically significant. The mean number of clicked target images of participants who focused their gaze mostly on the Centre of the screen is 4.97 ($stdev = 3.28$), and that of participants who focused their gaze mostly on the Bottom area is 5.21 ($stdev = 3.15$): $t(40) = -.24$, $p = .81$. Similarly, the mean of the number of seen target images of participants who focused their gaze mostly on the Centre of the screen is 7.5 ($stdev = 2.09$), and that of participants who focused their gaze mostly on the Bottom area was 7.94 ($stdev = 1.66$): $t(40) = -.76$, $p = .45$. Therefore, while grid speed and number of columns seem to have a clear connection with the most watched screen area (Top,

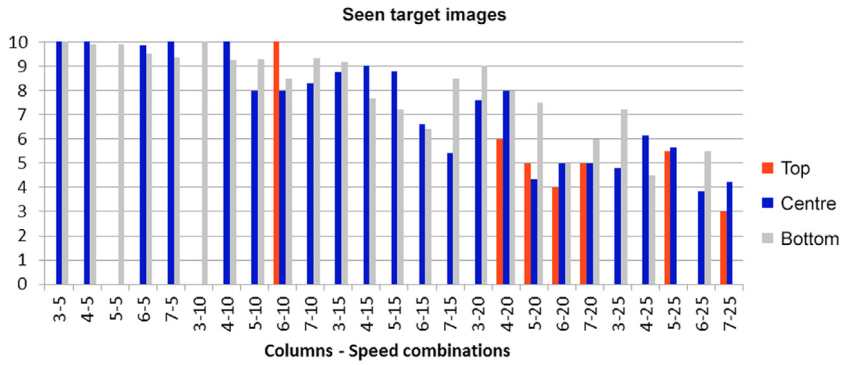

Fig. 8. Average number of target images seen by testers who mostly focused their gaze on the Top, Centre, and Bottom areas.

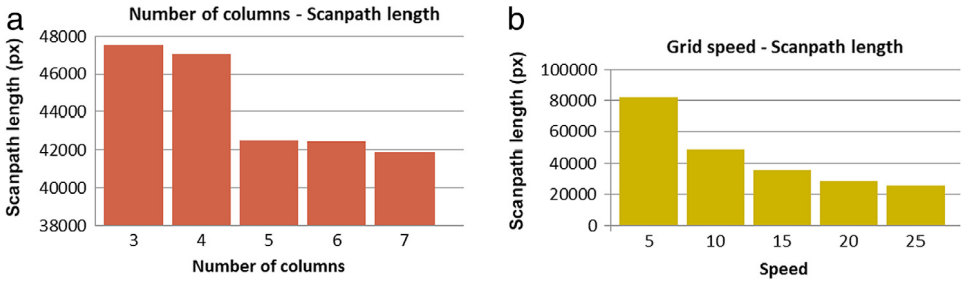

Fig. 9. Average scanpath length for the different numbers of columns (a) and speeds (b).

Table 1
Duration (in seconds) of each combination of number of columns and presentation speed.

		Number of columns				
		3	4	5	6	7
Speed	5	139	109	90	77	66
	10	70	54	45	39	32
	15	46	36	30	26	22
	20	35	26	22	20	17
	25	26	21	18	16	14

Centre, or Bottom), this behaviour does not directly affect user performance.

The average difference (in absolute value) between the percentages of fixations in the Bottom and in the Top areas (42.72%) is decidedly higher than that between the percentages of fixations in the Centre and Top areas (28.59%). The average difference between the percentages in Bottom and Centre (46.84%) is instead roughly similar to that between Bottom and Top (42.72%). In total, the percentage of fixations in the Bottom area is higher than that in Centre 126 times out of 249 (56.6%). Regarding the Top area, the percentage of fixations in Centre is higher than that in Top 236 times out of 249 (94.78%), while the percentage of fixations in Bottom is higher than that in Top 177 times out of 249 (71.08%).

4.3. Scanpath length

Another metric that can be considered is *scanpath length*, i.e. the sum (in pixels) of the Euclidean distances between all consecutive fixations. We observed that scanpath length decreases when grid speed and number of columns increase (Fig. 9), which, however, is quite obvious, since, with higher speeds or higher numbers of columns (and therefore more images), the test duration was shorter and testers looked at the screen for less time (Table 1 lists the durations of all presentations, expressed in seconds).

Therefore, we also considered a *normalized scanpath length*, computed by dividing the average scanpath length by the duration of the corresponding combination. As can be seen from the histograms in Fig. 10, the normalized scanpath length increases with number of columns and grid speed.

A Kruskal–Wallis test showed that there is a significant correlation between number of columns and normalized scanpath length ($Mdn = 1108.83$, $\chi^2 = 97.17$, $p < .001$). Moreover, there is also a significant correlation between speed and both normalized scanpath length ($Mdn = 1108.83$, $\chi^2 = 50.15$, $p < .001$) and scanpath length ($Mdn = 35556.82$, $\chi^2 = 190.73$, $p < .001$). Fig. 11 shows the scanpath and normalized scan path lengths depending on both number of columns and grid speed.

5. Discussion and conclusions

Unfortunately, it is not possible to definitely answer our first research question (“Is there any best combination of number of columns and scroll speed that guarantees a ‘good’ performance when the purpose is to find specific subjects within an image set?”) in an absolute way, because the best combination depends on the specific context. Of course, combinations with less columns and lower speeds were those characterized by the highest levels of accuracy (for both clicked and seen target images). However, these combinations were also little appreciated by testers, who found (for example) speed 5 (125 px/s) too slow and “boring”, as could be deduced from their informal comments. Moreover, testers found speed 25 (625 px/s) too fast and tiring, since a very high concentration effort was needed: in particular, the combination with the maximum number of columns (7) and the maximum speed (25) was judged as “almost impossible” by all testers. No particular subjective preference was instead expressed regarding the number of columns. We can notice that the average number of wrongly clicked images was always very low, which is likely due to the fact that target images were very well recognizable.

As a guideline for interaction designers, we can state that:

- When an accurate search is required, and the user must be able to click on the identified image, combinations with

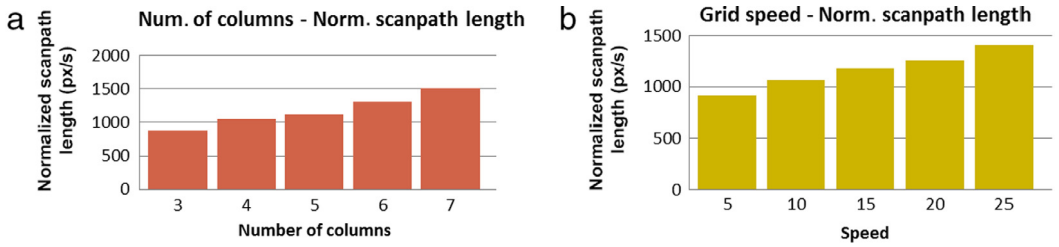

Fig. 10. Normalized average scanpath length for the different numbers of columns and speeds.

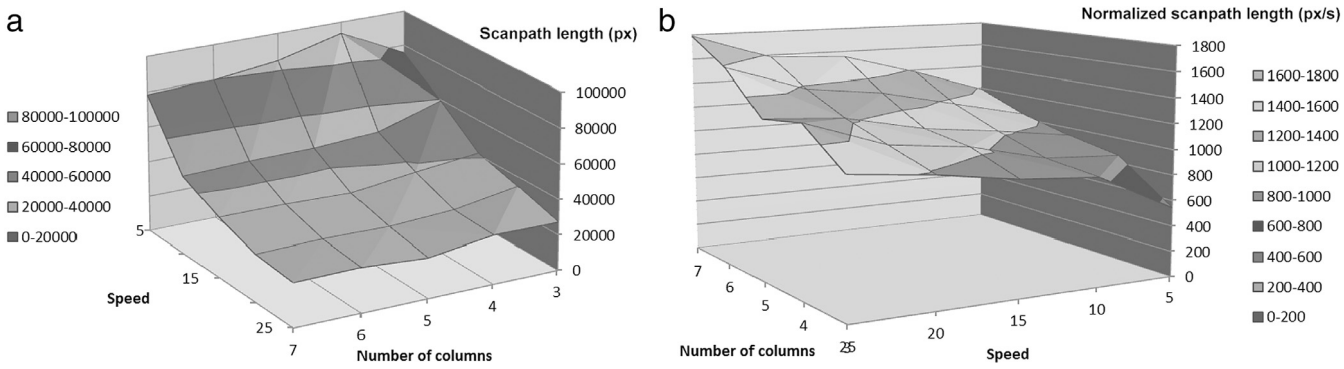

Fig. 11. Scanpath (a) and normalized scanpath (b) lengths depending on number of columns and grid speed.

a low number of columns and a limited speed should be chosen—approximately, in our experimental conditions, no more than five columns and speed 10 (250 px/s) at most. Alternatively, more columns and higher speeds could be selected (roughly up to six columns and speed 15), provided that the user can stop the presentation when he or she sees a target image (e.g. by pressing a key on the keyboard) to perform a precise inspection through a manual scroll.

- When an accurate search is not compulsory (because it is not necessary to search for all the target images in a set, like when simply seeking pictures for decorative purposes), combinations with a higher number of columns and greater speeds can be employed, so that the presentation time can be shorter. Depending on whether a click on target images is needed or not, moderate to high numbers of columns and speeds can be chosen (e.g. COLUMNS–SPEED combinations from 5–15 to 7–20). Speed 25 should however be possibly avoided, due to the negative experience reported by all testers.

Concerning our second research question (“Is image search characterized by any particular gaze behaviours? And, if so, do these behaviours depend on number of columns and scroll speed?”), we can certainly answer that yes, number of columns and scroll speed have an influence on the user’s gaze behaviour. In particular, with few columns and low speeds the user tends to look at the lower part of the screen, from where images enter. As the values of the two parameters increase, however, the gaze is shifted to the central screen area: since the search task becomes harder, the longer time available to identify target images during their route from Bottom to Centre provides the user with a certain “advantage”. The upper screen area is probably little watched because pictures are about to disappear and, “instinctively”, the user’s gaze is more focused on the new images that are arriving. Nevertheless, user performance (for both clicked and seen target images) seems not be clearly connected with the most observed area of the screen.

The average difference (in absolute value) between the percentages of fixations in the Bottom and Top areas is decidedly higher than that between the percentages in the Centre and Top areas, which may be due to the fact that when fixations are mostly concentrated in the central or upper parts of the screen, the gaze can more easily cross the border between the two regions. Approximately comparable to the average difference of the percentages of fixations between the Bottom and Top areas is instead that between the Bottom and Centre areas: this means that when the gaze is focused on the middle or on the lower region, it tends to stay there more steadily, without crossing the border between the two areas.

Bottom and Centre almost equally share the user’s attention. Comparing Centre and Bottom with Top, we can observe that the percentage of fixations in Centre is almost always higher than that in Top (94.78%), while the percentage of fixations in Bottom is higher than that in Top in fewer cases (71.08%).

It is also important to notice that the normalized scanpath length increases with grid speed and number of columns. This means that the “gaze span” in the time unit is higher for combinations with many columns and high speeds, which results in a more stressful experience for users. Another guideline for interaction designers is thus that the duration of presentations with many columns and high speeds should be limited, as the high normalized scanpath length makes the user’s searching task very demanding.

As examples of application scenarios for the automatic scrolling image grid, let us consider the following situations, characterized by different goals and needs:

- A physician needs to browse a vast database of images of skin tumours to compare them with that of a patient: to speed up the process, the automatic scrolling grid could be used, provided that a combination with a very high accuracy (close to 100%) is chosen, both for clicking and simple seeing. In this scenario, one of the first six combinations of Fig. 3 could be chosen (low number of columns and speed).
- In a clothing store, a screen may display a scrolling grid showing the different dresses available: once a customer has

seen a garment he or she likes, salespersons can be asked to show where to find the corresponding piece of clothes. In this case, users do not need to make mouse selections, and therefore higher speeds could be used. For instance, combinations 3–15, 3–20 or 4–20 could be selected.

- In an e-commerce website that sells clothes, a scrolling grid similar to that proposed in the previous example could be presented. In this case, however, users should also be able to select the garments they like, not just see them. In such a scenario, combinations 6–10 or 7–10 could be good choices.

Given the very limited research on image grid display, and especially on automatic scrolling grids, we think that this work can be the starting point for more thorough investigations on the subject.

Future studies will consider additional parameters (such as image size) and more values for controlled variables (e.g. number of columns). Moreover, variants of the grid layout will be taken into account, such as the "cylinder" display, in which images are arranged on the rotating surface of a 3D rendered horizontal cylinder. The mobile context will be considered as well, with the additional restrictions posed by limited screen sizes.

References

Beck, F., Burch, M., Vehlow, C., Diehl, S., Weiskopf, D., 2012. Rapid serial visual presentation in dynamic graph visualization. In: 2012 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC. IEEE, pp. 185–192.

Bederson, B.B., 2001. Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 71–80.

Cooper, K., de Bruijn, O., Spence, R., Witkowski, M., 2006. A comparison of static and moving presentation modes for image collections. In: Proceedings of the Working Conference on Advanced Visual Interfaces. ACM, pp. 381–388.

Corsato, S., Mosconi, M., Porta, M., 2008. An eye tracking approach to image search activities using RSVP display techniques. In: Proceedings of the Working Conference on Advanced Visual Interfaces. ACM, pp. 416–420.

Duchowski, A., 2007. Eye Tracking Methodology: Theory and Practice, vol. 373. Springer Science & Business Media.

Fan, X., Xie, X., Ma, W.-Y., Zhang, H.-J. and Zhou, H.-Q. 2003. Visual attention based image browsing on mobile devices. In: Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, vol. 1, IEEE, pp. I–53.

Igarashi, T., Hinckley, K., 2000. Speed-dependent automatic zooming for browsing large documents. In: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 139–148.

Kleiman, Y., Lanir, J., Danon, D., Felberbaum, Y., Cohen-Or, D., 2015. DynamicMaps: similarity-based browsing through a massive set of images. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, pp. 995–1004.

Liu, H., Xie, X., Tang, X., Li, Z.-W., Ma, W.-Y., 2004. Effective browsing of web image search results. In: Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval. ACM, pp. 84–90.

Oyekoya, O.K., Stentiford, F.W., 2006. Eye tracking—a new interface for visual exploration. BT Technol. J. 24 (3), 57–66.

Porta, M., 2006. Browsing large collections of images through unconventional visualization techniques. In: Proceedings of the Working Conference on Advanced Visual Interfaces. ACM, pp. 440–444.

Porta, M., 2009. New visualization modes for effective image presentation. Int. J. Image Graph. 9 (01), 27–49.

Ren, K., Sarvas, R., Calic, J., 2009. FreeEye: intuitive summarisation of photo collections. In: Proceedings of the 17th ACM International Conference on Multimedia. ACM, pp. 1127–1128.

Schaefer, G., 2010. A next generation browsing environment for large image repositories. Multimedia Tools Appl. 47 (1), 105–120.

Schoeffmann, K., Ahlström, D., 2012. Using a 3d cylindrical interface for image browsing to improve visual search performance. In: 2012 13th International Workshop on Image Analysis for Multimedia Interactive Services. IEEE, pp. 1–4.

Shneiderman, B., Kang, H., 2000. Direct annotation: a drag-and-drop strategy for labeling photos. In: Information Visualization, 2000. Proceedings. IEEE International Conference on. IEEE, pp. 88–95.

Spence, R., 2002. Rapid, serial and visual: a presentation technique with potential. Inf. Vis. 1 (1), 13–19.

Spence, R., Witkowski, M., 2013. Rapid Serial Visual Presentation: Design for Cognition. Springer.

Strong, G., Hoeber, O., Gong, M., 2010. Visual image browsing and exploration (Vibe): user evaluations of image search tasks. In: International Conference on Active Media Technology. Springer, pp. 424–435.

van der Corput, P., van Wijk, J.J., 2016. ICLIC: interactive categorization of large image collections. In: 2016 IEEE Pacific Visualization Symposium (PacificVis). IEEE, pp. 152–159.

Witkowski, M., Spence, R., 2012. Rapid serial visual presentation: An approach to design. Inf. Vis. 11 (4), 301–318.