Effects of Presentation Mode and Pace Control on Performance in Image Classification

Paul van der Corput and Jarke J. van Wijk

Fig. 1. How can we inspect a large set of images most comfortably, fastest, and with the least amount of errors? A: by viewing these by page *(static mode)*, or B: by putting them on a conveyor belt and let the images pass by *(moving mode)*?

Abstract—A common task in visualization is to quickly find interesting items in large sets. When appropriate metadata is missing, automatic queries are impossible and users have to inspect all elements visually. We compared two fundamentally different, but obvious display modes for this task and investigated the difference with respect to effectiveness, efficiency, and satisfaction. The *static mode* is based on the page metaphor and presents successive pages with a static grid of items. The *moving mode* is based on the conveyor belt metaphor and lets a grid of items slide though the screen in a continuous flow. In our evaluation, we applied both modes to the common task of browsing images. We performed two experiments where 18 participants had to search for certain target images in a large image collection. The number of shown images per second (pace) was predefined in the first experiment, and under user control in the second one. We conclude that at a fixed pace, the mode has no significant impact on the recall. The perceived pace is generally slower for moving mode, which causes users to systematically choose for a faster real pace than in static mode at the cost of recall, keeping the average number of target images found per second equal for both modes.

Index Terms—RSVP, image classification, image browsing, multimedia visualization

1 Introduction

Visualization provides a powerful way to get insight in vast amounts of data. Good visualizations enable users to quickly find the pattern or item they are looking for. When designing a visualization, it sometimes happens that a particular part of the data is hard or impossible to query, for example when appropriate metadata is missing. Suppose that the task is to look for a person, geographic location, or cell structure, but the exact name or properties are unknown; or the task is to look for interesting elements, not having a particular target in mind. In these cases, it is impossible to guide the user directly to this information, and the only solution is to present all potentially relevant items and let the user decide what is interesting.

In this paper we focus on an instance of this problem that almost any computer user is familiar with: image browsing and search. This is an ubiquitous task in both private and business applications, for instance browsing a personal image collection, searching for an image on the web, scanning security footage, and filtering inappropriate content on web services. These tasks usually involve the inspection of thousands of images. While machines are still becoming better at pointing out images and features of interest, for example in the domain of face detection and number plate recognition, in most cases we still have to rely on human judgement. This results in long and demanding sessions, and therefore we are interested in optimizing the human performance in such tasks. Obviously, one wants to find as many interesting results as possible in the shortest amount of time. In this paper we compare and investigate the effects on human performance of two similar looking, but fundamentally different presentation modes as depicted in Figure 1. The first mode is a static page that contains a grid of (thumbnail) images and is refreshed at once. The second mode uses the *conveyor belt metaphor*, and constantly slides the image grid from one side of the screen to the other.

A comparison of the commonly used page metaphor with the conveyor belt metaphor is relevant, because a conveyor belt looks natural in this situation but is not much applied in practice. The main advantage over the page metaphor is that the information flow is continuous, and there is no abrupt screen refresh. For measuring the effects of both modes, we look at usability goals such as described by Nielsen [11] and the ISO 9241-11 standard [8]: *effectiveness*, *efficiency* and *satisfaction*. It is not easy to optimize all three goals simultaneously. When for example the number of presented items per second increases, the user is put under more stress and will miss more targets.

In order to find out how effectiveness, efficiency, and satisfaction are interrelated in the task of browsing images, we conducted some experiments. The results can be used to determine which configuration suits the task best. Sometimes, the effectiveness is of utmost importance, like in security applications where terrorists should not be missed, but when the goal is to acquire just a global overview of a dataset, it is sufficient to make an quick scan. Finally, if tasks have to be done frequently or for longer periods of time, user satisfaction plays an important role. In the end, it is clear that we want to optimize the overall usability with the emphasis on one or two aspects.

Image and text recognition at high paces have been investigated in the past, and we discuss this in Section 2. We see that the use of the conveyor belt metaphor has not yet been considered and evaluated, how this can be done is analyzed in Section 3. Section 4 contains a detailed description of the experiments we conducted, and in Sections 5 and 6 we present the results. Finally, in Sections 7 and 8 we discuss these results and give our conclusions.

- *Paul van der Corput is with Eindhoven University of Technology. E-mail: p.n.a.v.d.corput@tue.nl.*
- *Jarke J. van Wijk is with Eindhoven University of Technology. E-mail: j.j.v.wijk@tue.nl.*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014. Date of publication 11 Aug. 2014; date of current version 9 Nov. 2014.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier 10.1109/TVCG.2014.2346437

1077-2626 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

2 Related work

The recognition of visual items at high speeds has been studied for a long time. Since the late 1960s, experiments have been conducted to investigate how sequences of visual stimuli are perceived by humans. The art of presenting these sequences rapidly was called *Rapid Serial Visual Presentation* (RSVP) by Forster [5]. Forster used this technique to study the effects of sentence complexity. He put individual words on successive frames of a 16-mm film and projected this movie at 16 words per second.

Later, Spence [14] noted that RSVP has a variety of interesting applications such as searching for interesting books or movies. Since then, various parameters have been explored in order to maximize the information transfer to the user, i.e., increasing the performance. A comprehensive overview of all these aspects is given by Spence and Witkowski [15]. Spence et al. [13] explored the space-time trade-off, where one has to choose between showing one large image for a short time period, and showing many smaller images for longer time period. They found that presenting smaller images simultaneously is less error prone than presenting large images one-by-one. Cooper et al. [3] experimented with several static and moving RSVP modes, and the effect on identification rates and user acceptance. Their conclusion is that static modes are in general better than animated modes with respect to recognition success as well as user preference. In animated modes, the use of overlapping images reduces the speed at which images have to move, which can make it easier to track them. Brinded et al. [1] showed that small overlap has little effect on the error rate, perceived speed, confidence, and difficulty and can therefore in some cases be acceptable.

Viewing images at a high pace is relevant for applications in several areas. An example is content based image retrieval, where human input is vital. Hauptmann et al. [6] and Luan et al. [9] used RSVP techniques in their application to be able to search for images at record speeds. The possibilities of RSVP have also been explored in order to rapidly scan terrain data. Mardell et al. [10] considered a static and an animated mode for use in search and rescue missions, and found that the static mode is the most effective, although users did not have a clear preference.

When it comes to browsing large image collections, commonly used approaches are showing images one-by-one, and showing a grid of thumbnails. Some research has been done on unconventional presentation modes, for example by Porta [12]. Porta presents the elastic image browser, which is based on a conventional grid and can be used to quickly get an idea of the collection's content. Christmann et al. [2] investigated some dynamic three dimensional visualizations and the effects of geometrical distortions on effectiveness, efficiency, and comfort. They conclude that perspective views have the potential to improve visual search in unstructured image collections. If selection of interesting items is necessary, the conventional solution is using the point-and-click technique, but Corsato et al. [4] showed that using eye tracking systems could be an effective alternative for this.

3 Problem analysis

The related work shows that much is already known about the human capabilities with respect to rapid visual stimuli perception. Differences between static and animated modes have been investigated, but there are still questions left. The moving image grid as presented in the introduction has for example not yet been compared to a static grid. Furthermore, longer experiments are needed to investigate the effects on usability over time, because the task can be very demanding. Finally, the effect of pace control has not been investigated.

3.1 Evaluation of usability

The usability of a system can be decomposed into effectiveness, efficiency, and satisfaction. Below we describe how these aspects can be evaluated for image presentation modes. We consider image retrieval as our main task, and consider cases where about 1 out of 100 images have to be identified as target images.

3.1.1 Effectiveness

The first goal is to find as many target images as possible, but at the same time not to mark irrelevant images. In image retrieval, the metrics *precision* and *recall* are used. To measure these, we have to keep track of the number of target images found (*tp*, true positives), the number of target images missed (*fn*, false negatives), and the number of falsely marked target images (*fp*, false positives). The occurrence of identifying a non-target image (*tn*, true negative) is in this case not so relevant.

The precision is the ratio between the targets marked correctly and all marks, and can be computed as $tp/(tp+fp)$. The recall is the ratio between found target images and presented target images, or $tp/(tp+fn)$. We call a presentation mode effective when users produce few false positives and false negatives, hence resulting in high precision and recall. Which of the two metrics is most important depends on the application, in this paper we therefore consider both in our analysis.

3.1.2 Efficiency

The second goal is to scan through many images in the shortest amount of time. We therefore use the measure *pace*, which is defined as the average number of images that enter the display area per second.

3.1.3 Satisfaction

The third goal is to give the user an as comfortable experience as possible. One could perform physiological measurements, a simpler alternative is to consider user responses, such as preference (what do users like and why?), confidence (how effective do users think they are?), and perceived stress (how demanding is the task?).

3.2 Presentation modes

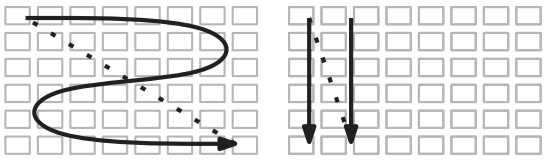
We compare two image presentation modes. One is based on the page metaphor, which allows the user to go to the next page of images once all images have been inspected. The images have a fixed position on the screen and they (dis)appear simultaneously on each page turnover. The second metaphor is that of the conveyor belt, where there is a constant flow of images from one side to the other side of the screen. For the task of observing and searching images this seems a very natural metaphor. Similar work is done in factories where products have to be inspected, for instance the process of removing broken bottles from a conveyor belt.

Since the essential difference between the two metaphors is whether images are moving, we call in the remainder of this document the mode based on the page metaphor *static mode*, and the mode based on the conveyor belt metaphor *moving mode*. In order to make a fair comparison, both modes show an image grid with the same dimensions. For our experiments, we use the dimensions used by Cooper et al. [3], who used 6 rows and 8 columns their tile mode. The static mode presents consecutively 8 columns of this grid, while the moving mode slides this grid continuously over the screen from left to right. We have chosen for this direction and not from top to bottom, because it best matches the situation where a factory worker stands next to a conveyor belt and sees items for inspection coming from left to right.

3.3 Hypotheses

The user's search pattern depends on the mode. In static mode, each page has to be scanned entirely which induces more or less random eye movements. When searching in moving mode, one can take advantage of the automatic displacement of the images such that the focus does not need to be distributed equally over the screen area. The user can focus on the left side of the screen inspecting the stream column by column as illustrated by Figure 2. Since the conveyor belt metaphor also seems natural and is widely applied in factories, we expect the moving mode to be beneficial for efficiency and user satisfaction. Based on these assumptions, we use the following hypotheses for our experiments:

H1 *The moving mode enables users to achieve a higher recognition accuracy than the static mode, in case the pace is fixed.* The natural looking constant flow of a conveyor belt gives users the



(a) Static mode (b) Moving mode

Fig. 2. Obvious search patterns in both modes.

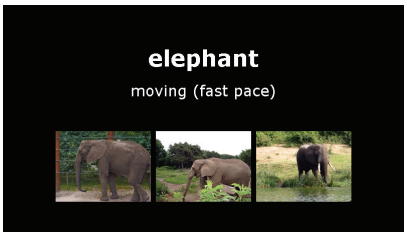


Fig. 3. Example of an overview screen that was shown for six seconds preceding to each sequence. The category is expressed textually, and for extra clarity, also with three sample images from the target category. Sample images were not used in the rest of the sequence.

opportunity to use more structured search patterns than the static mode does. In static mode it is less clear when images disappear, and hence the ideal search speed is hard to determine.

H2 *Users prefer the moving mode to the static mode.* The same reason as above applies here, and hence this effect should get stronger if the pace increases. At some pace, the transit time is so short that scanning the entire matrix is impossible.

H3 *When they can control the pace, users are more productive in moving mode than in static mode.* It is tempting to carefully inspect all images in static mode, which takes a lot of time and is not needed in most cases. In moving mode, the users have to press a button to lower the pace, which makes them conscious of the decrease in pace.

H4 *In both modes, users will eventually suffer from the fatigue effect.* Image analysis is a demanding task. In the introduction, we gave some examples of tasks that require the inspection of thousands of images. It seems plausible that human focus will drop along the way, regardless of the presentation mode.

4 Experimental Design

In order to investigate the effects of presentation mode and user controlled pace on the user's performance we designed two experiments. The experiments consisted of several image sequences and subjects had to recognize targets in these. In both experiments, half of the sequences were presented in static mode, the other half in moving mode. In the first experiment, we chose three fixed paces for the sequences and measured how pace and mode affect accuracy and user comfort. In the second experiment, we enabled the users to select a comfortable pace, and measured how it affects the productivity and accuracy. In an ideal situation we could fix the accuracy and measure its effect on the pace and user comfort in a third experiment. However, with humans as test subjects, this is impossible.

4.1 Task

In both experiments, the participants had to recognize images originating from a given *target category*, which were pseudo-randomly distributed over large sequences of images. The participants were instructed to press the SPACE bar on the keyboard for every target image that entered the screen, e.g., three keystrokes were required if three targets were visible. The allowed response time equaled the *image transit time*, which is the time that the image is visible on screen. In Experiment 2, additional buttons were available to control the pace.

No information about the number and distribution of target images was provided to the participants, and they also had to come up with their own search strategy, as no hints were given about efficient patterns. The participants were instructed to try to make as few mistakes as possible, and to motivate them, we awarded the fastest and most accurate participant with the *Ultimate Image Classification Award* and a 25-euro gift card. The rules of this game were that each mistake counts for a penalty of 12 seconds, and that the one with the shortest execution time including penalties is the winner.

4.2 Participants

Both experiments were performed by 18 participants, 5 female, 13 male with ages between 20 and 62 years. The backgrounds of the participants are diverse: both people with a computer science (11) and non-computer science (7) education participated. Other skills that can be relevant for the results are experience with image annotation and computer games. Four of the participants are image classification experts, and all participants use computers on a regular basis.

4.3 Sequences

All sequences were constructed as follows. For the first 6 seconds, an overview screen was shown with the target category, mode, and in Experiment 1 also the pace. Figure 3 gives an example of such a screen. After the overview screen, an empty black screen was shown for 2 seconds. Finally, the images were shown according to the presentation mode, category, and pace. The target images were pseudo-randomly distributed over the sequence with the constraint that there was always at least one column separating the target images to avoid high local densities. We aimed to avoid target clusters, because these could potentially make the task harder.

4.4 Images and categories

For a realistic setting, we used the MIRFLICKR-25000 image collection [7] as baseline. This collection contains a variety of 25,000 images from the Flickr website. An impression of the diversity of this collection is given in Figure 1. During the experiments, these are the non-target images, which we refer to as *noise*. In addition, we made an image collection by hand consisting of 16 familiar target categories: cars, airplanes, busses, elephants, tigers, bears, giraffes, tennis (rackets), cycling (races), basketball (baskets), rowing (boat races), apples, bananas, oranges, kiwis, and pineapples. These images also come from Flickr, but not necessarily from the MIRFLICKR collection. Each of the categories contains about 30 images, which is enough to ensure that each image is shown at most once per participant.

The images were chosen carefully to ensure that there could be no doubt that they belong to their category. Basic constraints were that the target was on the foreground, not too much obscured, and not in a strange configuration. Furthermore, we filtered the noise collection manually in several passes for images belonging to target categories or that could otherwise potentially be classified as targets. In the end, for each of the displayed images, it should be immediately clear whether it is a target image or not.

4.5 Factors

Below is a summary of all the factors under investigation in both experiments. In Experiment 1 we had three parameters: mode, pace, and round. The modes under investigation were static and moving mode as described in Section 3. We chose for three fixed paces, namely: 8, 12, and 16 images per second. A pilot study revealed that in this configuration, human mistakes are exceptions at the slowest pace and are unavoidable at the highest pace. The third parameter is used to measure the learning effect, i.e., whether the participants get better at the task during the experiment. Therefore, each session with all combinations of mode and pace is repeated such that the results of round 1 and round 2 can be compared. In Experiment 2 we had the same two modes, however, the pace was controlled by the user and the learning effect is there measured along the sequence.

Fig. 4. Examples of the images of the 16 categories that have been used in the experiments, with the average precision p and recall r observed in Experiment 1.

4.6 Hardware and implementation

A Lenovo W520 laptop (Intel Core i7 2630QM CPU, Nvidia Quadro 1000M GPU, 4GB RAM, Windows 7) has been used for all experiments. The use of a laptop makes it possible to travel with the same experimental equipment towards participants, which makes is easier to collect data from a diverse group. The laptop has a 15.6 inch LCD display with a resolution of 1600×900 pixels. Participants were allowed to choose their preferred distance to the screen, which was in each case approximately 1.5 feet. The display's refresh rate was set to 60Hz, and was equal to the frame rate produced by the software. This software was written in Java using standard Java graphics libraries. All images were pre-loaded in memory before the start of each sequence to eliminate glitches due to slow disk reads.

In all experiments we used the same image grid size, namely 8 columns wide and 6 rows high, so 48 images per page. Each image was cropped to the same aspect ratio and scaled to 190×140 pixels. The space between images was 10 pixels, and the background color was black. The number of columns entering the screen in moving mode is the pace divided by 6. Since the total column width is 200 pixels, the horizontal displacement in moving mode is $200 \times pace/6$ pixels per second. This comes down to 267, 400, and 533 pix/s for the three paces being tested in Experiment 1. An illustration of this configuration is given in Figure 1, and it shows how the screen area (colored rectangles) is translated over the image grid.

4.7 Procedure

The procedure per participant was as follows. First, the participant got instructions about the presentation modes and task. Next, he or she performed four training sequences to check whether the task was clear and to let the participant get used to the modes and paces. After these sequences, the first experiment began, directly followed by the second experiment. The participants performed this procedure one-by-one, while the instructor was sitting next to, and slightly behind the participant to minimize the interference. The steps of the procedure are detailed out below.

4.7.1 Instructions and training sequences

At the start of the procedure, the static and moving modes were explained and the task was presented. The participants were told that the procedure would take approximately 30 minutes, and that a price is awarded for the participant with the fewest mistakes and fastest time. After the participant agreed that everything was clear, four test sequences were run, for which the slowest and fastest pace for both modes of Experiment 1 were used. The main reason for these training sequences was to reduce the learning effect in the two experiments, because the learning curve is the steepest in the beginning. In these training sequences, the category was fixed (flowers). No data was logged during these sequences, allowing the instructor to point out mistakes to make sure that the task was fully understood by the participant before the actual experiment began.

4.7.2 Experiment 1

The first experiment was focused on the effects of pace and mode on the number of errors and user comfort. This was done by presenting six sequences with all combinations of mode and pace (in randomized order to deal with any learning effect). A second session of six again randomized combinations was presented to enable us to measure this learning effect. This resulted in 12 sequences in Experiment 1, each consisting of 480 images. Each time, five images were drawn from the target category, giving the sequence a target density of 1/96, or one target per two pages on average. A random category was assigned to each of those 12 sequences to make sure that participants do not familiarize themselves with one specific category during the experiment. We are aware that variations in recognition difficulty for images within and between categories are unavoidable. We therefore distributed random subsets of target categories randomly over the sequences. Figure 4 depicts representative examples of these categories and an indication of their average difficulty.

After each sequence, the participant had to answer two questions: (1) "How comfortable was the pace?" and (2) "How many targets do you think you have missed?" For both questions, a score on a scale from 1 to 5 could be assigned. At the end of Experiment 1, the participant could give his or her preference for either the static or moving mode.

4.7.3 Experiment 2

In the second experiment our focus was on practical applications, where user comfort plays an important role. The setup was similar to Experiment 1, except that subjects were now enabled to adjust the pace during the experiment. In static mode, the ENTER button was used to move to the next page. In moving mode, the LEFT ARROW and RIGHT ARROW buttons could be used to respectively decrease and increase the pace by one image per second. As a result, we were able to measure the development of preferred pace during the sequence and the corresponding number of mistakes. Experiment 2 consisted of two sequences: first moving mode and then static mode. Because comfort can be measured better in longer sessions, we extended the sequence size to 2400 images (5 times as long as the sequences in Experiment 1). By using such long sessions, there is no need to repeat them. The learning effect (if there still is any after so many sequences) can be measured along the sequence. Besides the addition of user control, no adjustments were made to the task and the density and distribution of the targets. At the end of both sequences, we asked the participants for their confidence on the number of correctly spotted images. After Experiment 2 there was a short questionnaire where participants were asked for their preference with respect to mode and user control.

5 Results Experiment 1

In the first experiment, the pace was fixed per sequence. Below we present the results on precision, recall, response time, and user feedback; and compare between static and moving mode.

5.1 The effects of mode and pace on precision and recall

Figure 5a depicts the effects of mode and pace on the recall. It appears that an increase of pace generally results in a decrease of recall. The choice between static and moving presentation mode does not seem to have much impact. The precision seems not to be affected by pace or mode. This is confirmed by Table 1, which presents a three-way

(a) Pace and mode versus recall. (b) Pace and mode versus precision. (c) The fatigue effect per pace. (d) The fatigue effect per mode.

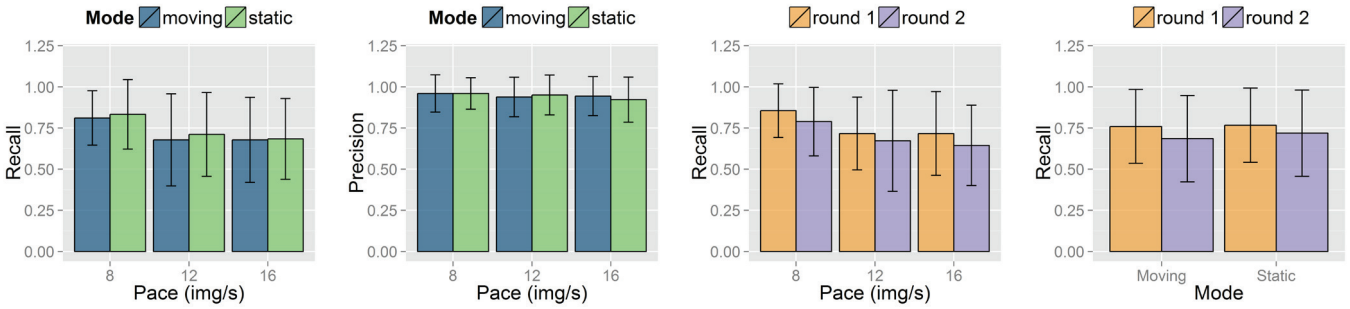


Fig. 5. Figures (a, b) show the effects of mode and pace on recall and precision. The learning (or fatigue) effect can be measured as the change in recall between the beginning and the end of the experiment, i.e., round 1 and 2 as depicted in Figures (c, d). All bars represent averages and the error bars indicate the standard deviations.

ANOVA for presentation mode (static, moving), pace (8, 12, 16 images/s), and round (first, second session). The ANOVA tells us that none of the factors influences the precision, and that pace has a significant effect on recall.

Table 1. Three-way ANOVA for the effects of mode, pace, round, and their interactions on precision, recall, and average response time. Indicated are the degrees of freedom (df), F-values (F), and p-values (p). Significant p-values are emphasized.

Factors	df	precision		recall		response time	
		F	p	F	p	F	p
Mode	1	0.05	0.83	0.39	0.53	3.22	0.07
Pace	2	0.92	0.40	7.64	***0.00***	29.4	***0.00***
Round	1	0.56	0.45	3.51	0.06	0.94	0.33
Mode : Pace	2	0.36	0.70	0.06	0.94	0.65	0.52
Mode : Round	1	0.68	0.41	0.16	0.69	3.81	*0.05*
Pace : Round	2	0.11	0.89	0.07	0.93	0.23	0.80
M : P : R	2	0.09	0.92	0.35	0.70	2.75	0.07
Residuals	204						

5.2 The fatigue effect

The participants had to perform two sets of six sequences, which we call Round 1 and Round 2. This enables us to measure the number of mistakes made per round, and so the learning effect (or fatigue effect, in case the participant's performance drops). Figure 5c shows that the latter seems to be dominating. Furthermore, there is no strong difference between modes, see Figure 5d. This is also supported by the ANOVA in Table 1, as the effect of round on recall is not significant.

5.3 The response time

The *response time* is the time between the event that a target image enters the visible screen area and the event that the participant presses the SPACE bar. The analysis of the response time can give insight into how close the participants are to missing an image, i.e., when the response time exceeds the *transit time* (time that an image is visible). At a pace of 8 images per second, this transit time is $48/8 = 6$ seconds. For 12 and 16 images per second, this is 4 and 3 seconds respectively. In moving mode, this is slightly more because images can be displayed partially. Figure 6 shows that targets are generally identified earlier in moving mode, and well within the required time frame. The ANOVA in Table 1 partially confirms the findings from the visual inspection; the pace has a significant effect on the response time but the mode has not, possibly because of the large variance. In case of a missed target, the response time is infinite, hence for the analysis of response time, we did not take false negatives and false positives into account.

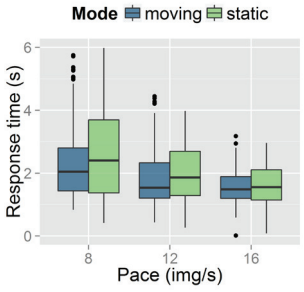


Fig. 6. The effect of mode and pace on response time.

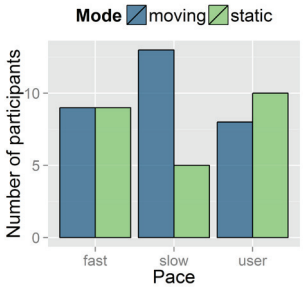


Fig. 7. User preference for the mode based on the pace: slow, fast (Experiment 1), and user defined (Experiment 2).

5.4 The perceived pace and recall

After each sequence, the participants were asked how they perceived the pace and to guess how many targets they had missed. We can check how these subjective responses correspond to the real pace and number of mistakes. Figures 9a and 9b show the relation between real and perceived pace for the static and moving mode, whereas the relation between real and perceived misses is shown in Figures 9c and 9d. Not surprisingly, the real pace and misses influence the perceived pace and misses. Participants were able to see differences in pace, and felt to make more mistakes whenever they actually did so. This can be observed as the disks being large on the diagonal. A more interesting effect is the shift in perceived pace between the modes: static mode is generally perceived faster than moving mode. We performed two two-way ANOVAs to measure (1) the effects of real pace and mode on the perceived pace, and (2) the effect of real misses and mode on the number of perceived misses. The mode appears to have a significant effect on perceived pace $F(1,210) = 12.6, p = 0.00$; this is also (but slightly less) the case for the perceived misses $F(1,204) = 4.25, p = 0.04$. The effect of mode was significant (p-values are far below 0.01) as expected.

5.5 User's opinion and preference

After both experiments, participants were asked for their preferred mode. Since the pace was varied in Experiment 1, a distinction is made between working at a slow and fast pace. Figure 7 shows that moving mode was preferred when a slow pace was presented, and that opinions were divided with regard to a fast pace. Participants were also asked to motivate their preference. Like their preference, the responses to this open ended question were also very diverse and even

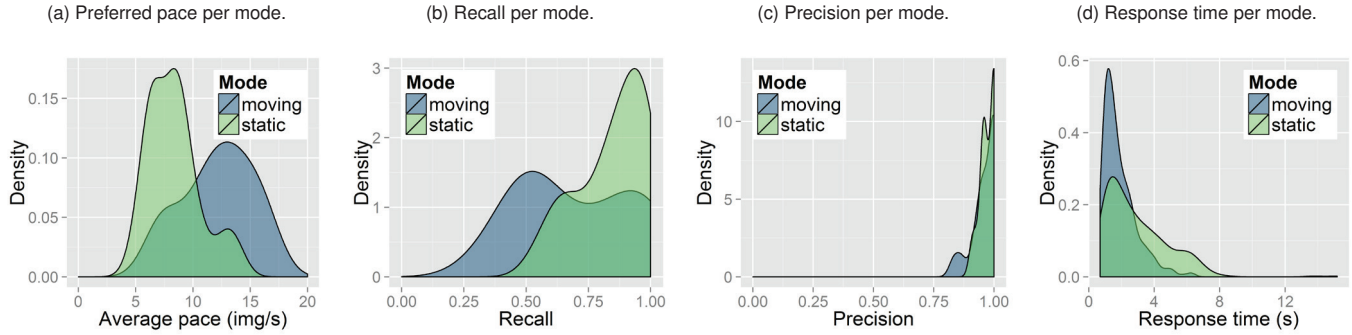


Fig. 8. The differences between modes when participants can control the pace shown by means of density plots. The higher the density, the more participants that show the corresponding value on the x-axis. The area under each curve always equals 1.

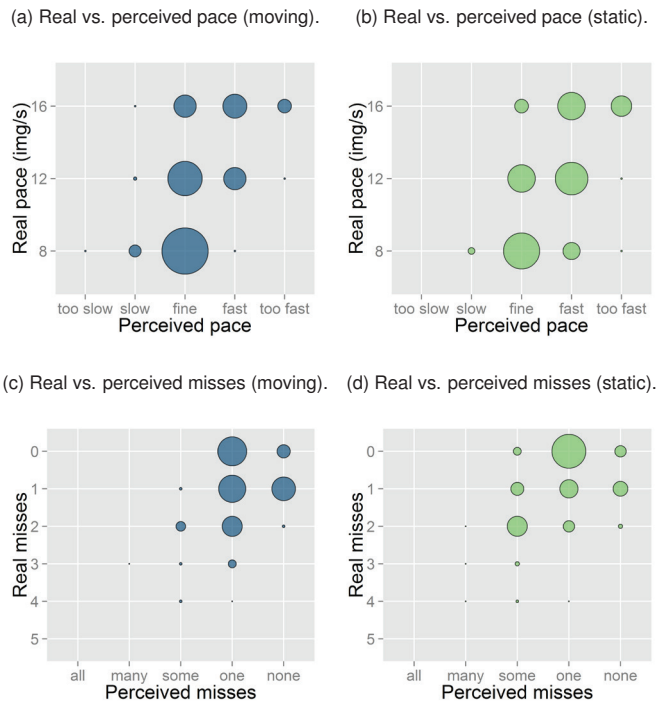


Fig. 9. The relation between real and perceived pace and number of misses in Experiment 1. The number of responses is represented by the disk's size.

contradicting. Some participants found the moving mode comfortable, while others said that it causes headaches:

- Participants who preferred the moving mode mentioned that it provided them a structured way of searching, namely by continuously scanning vertically. They reported that this pattern is easier than scanning from left to right and top to bottom, and therefore reducing the chance of missing targets. By scanning this way, their eyes were always focussed on the same part of the screen. This required less eye movement and hence led to a more comfortable feeling. Furthermore, some participants found it difficult to use the static mode, because of the abrupt page refresh and therefore preferred the moving mode's natural flow.
- Participants who preferred the static mode noted that the moving mode causes motion blur, which is annoying and might lead to mistakes. Some participants found it tiresome to keep track of moving images, and reported tired eyes and felt that this mode could potentially cause headaches. The static mode was also felt to be less frantic. One participant thought of a psychological effect caused by the static mode that forced him to scan the page quickly before the page is refreshed, and therefore increasing the attention.

6 Results Experiment 2

In the second experiment, the participants were enabled to control the pace. Below we present the result of similar measurements as in Experiment 1 and the deviations from these caused by pace control. Furthermore, we investigate the ranges of paces that were preferred by the participants.

6.1 Preferred pace and its effect on precision and recall

Figure 8a shows a density plot with the distribution of preferred paces in both modes. The preferred paces are calculated by dividing the number of images per sequence (which was 2400 in each case) by the sequence duration. Participants used a faster pace when in moving mode. The corresponding impact on recall and precision is shown in Figures 8b and 8c. It appears that while static mode results in a lower preferred pace, it does result in a higher recall, and the opposite holds for moving mode. The mode seems to have little effect on the precision. Another aspect is the response time, which is plotted in Figure 8d. In moving mode, participants detected images within 3 seconds in almost all cases. In static mode, such reaction time is common but a significant amount of the responses take place between 3 and 8 seconds, with some outliers between 8 and 16 seconds. Table 2 shows that in this experiment, the presentation mode has a significant effect on average pace, recall, and response time. An important note here is that the variance between participants was considerable, as can be seen from Figure 8. Between the fastest and the slowest pace, there was for example a difference of a factor 3, and the recall ranged from 0.32 to 1. Only the precision was consistent. The actual standard deviations are shown in Table 2.

Table 2. The effects of mode on several metrics in Experiment 2, significant values are emphasized. For both modes, the mean and standard deviation are shown. A Welch Two Sample t-test was used for response time, and paired t-tests were used for the other metrics. The fourth column shows 95% confidence intervals on the difference between moving and static mode.

Metric	Mean, st.dev.		t-test	
	Moving	Static	95% conf. int.	p-value
Average pace	12.1, 3.08	8.37, 2.27	(2.06, 5.50)	***0.00***
Precision	0.96, 0.05	0.97, 0.03	(-0.04, 0.02)	0.42
Recall	0.69, 0.23	0.85, 0.14	(-0.25, -0.06)	***0.00***
Response time	1.89, 1.02	2.94, 1.94	(-1.27, -0.82)	***0.00***
Hits per second	0.09, 0.04	0.07, 0.02	(-0.01, 0.03)	0.15

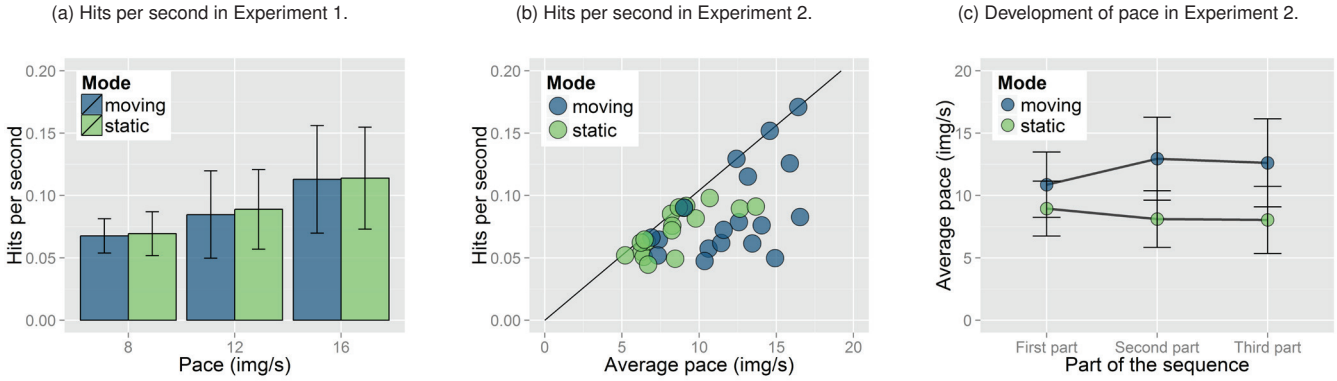


Fig. 10. Figure (a, b) show the relation between pace, mode, and hits per second in both experiments. A breakdown of the pace development in three stages in Experiment 2 is shown in Figure (c). The error bars in Figure (a, c) indicate standard deviations.

6.2 Hits per second

Since the increase of pace in moving mode is accompanied by a decrease in recall, it is interesting to check whether there is a difference in average number of true positives per second. We call this metric *hits per second*, and calculate it as

$$\frac{\#\text{ true positives} \times \text{average pace}}{\#\text{ images in the sequence}}. \tag{1}$$

Figure 10a and 10b show the hits per second for both modes in Experiment 1 and 2. The black diagonal in Figure 10b indicates the maximum number of hits per second for a given pace and the target density used in the experiment. We see that for a fixed pace there is hardly any difference between the modes. Furthermore, we observe that an increase in pace leads to an increase of found targets per second. This pattern is maintained at least until 16 images per second.

6.3 The fatigue effect

We can check for global pace changes within the sequences, which can indicate the severity of the fatigue effect. In Figure 10c, the sequences are broken up into the first, second, and third part of the presented images. It shows the average preferred pace in each part of the sequences, where a distinction is made between the two modes. The difference in preferred pace between static and moving mode is again clear, but apparently the modes also influence the pace development within the sequences. A two-way ANOVA with mode and sequence part as independent variables, and pace as dependent supports this observation: $F(1,102) = 48.6, p = 0.00$ for the effect of mode; $F(2,102) = 0.45, p = 0.64$ for the effect of round, which is not significant; and $F(2,102) = 2.97, p = 0.06$ for their interaction.

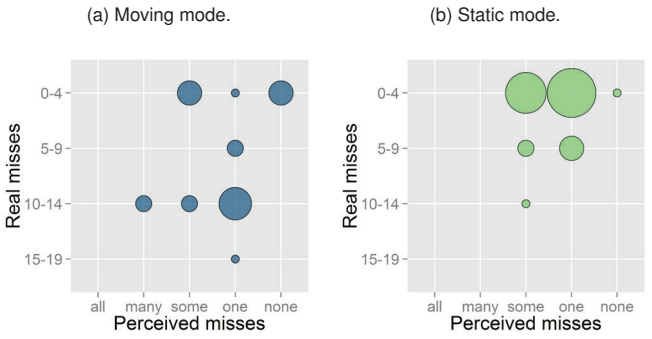


Fig. 11. Relation between real and perceived number of misses in Experiment 2. The number of misses are combined in groups of five.

6.4 User response and preference

Figure 7 shows that no mode is a clear winner as to the user preference in Experiment 2. We see a similar pattern in the reactions as after Experiment 1, with very diverse responses:

- Participants who preferred the moving mode found this mode less stressful because the up-down scanning method results in less eye-movement. In static mode, scanning feels more chaotic and the technique tends to change between pages. One participant noted that it is tempting to take too much time assessing all images in static mode, while movement forces progress.
- Participants who preferred the static mode felt to be more in control in static than in moving mode. They furthermore said to feel less pressure on their eyes due to the absence of motion (blur), and therefore call it less tiresome. Finally, one participant found it easier to scan from left to right rather than up and down, something that is difficult in moving mode.

The user control was very much appreciated in Experiment 2. All participants indicated to like it, mainly because of the ability to slow down to inspect the more difficult cases or when the level of attention becomes weak, making it a more comfortable experience. After the experiments they said to be more confident on their score, which is however not directly confirmed by their response on perceived misses, see Figure 11. Only four participants felt to have made no mistakes, and furthermore the confidence does not seem to have improved since the first experiment. Just as with the user preference, it is not so clear which mode maximizes the confidence (in other words, minimizes the perceived number of misses).

After Experiment 2, we asked the participants for their mode of preference in case they had to do this type of work for the entire day. The result was that eight participants supported the moving mode, also eight were in favor of the static mode, and two did not have a clear preference. The reasons brought by them in favor of moving mode were that it felt calmer by the reduction of eye movement, because you do not have to scan each corner of the screen like in static mode. They said that this mode made it easier to keep their concentration. The natural flow has the advantage that no refocus is needed like it is the case after a page refresh in static mode. Two participants commented however that moving mode is only preferred when the pace can be controlled and motion blur is reduced. Reasons that supported the static mode were that, with pace control, it is very easy to take breaks which makes it less tiresome and results in less mistakes. One participant found that after looking at moving mode for a long time, there is some after effect where everything else starts to move.

7 Discussion

Below we relate the results to the hypotheses and discuss the possibilities for future experiments.

7.1 Analysis of the results

The statistics obtained from Experiment 1 tell us that when the pace is kept constant, there is no significant difference between presentation modes with regard to precision and accuracy. Therefore, we can reject Hypothesis 1, i.e., the ability to scan a smaller part of the screen does not make it easier to absorb visual information. Interesting to mention is that in contrast to the conclusions of Cooper et al. [3], we observed no significant degradation of recognition success and user acceptance when switching from static to moving mode. Also, our implementation of the moving mode does not support the advised *capture effect* where images have a fixed location for at least 100 to 200 milliseconds. We think that this difference in results is caused by showing n (in our case six) rows at the same time instead of a single stream as used by Cooper et al. The horizontal speed of the images can therefore be reduced with a factor n if the pace and image dimensions are kept constant, and this image speed plays a crucial role in the recall as has been investigated by Brinded et al. [1].

We see that in Experiment 1 the moving presentation mode is perceived slower than the static mode at the same pace. Furthermore, the participants seem to be more confident when using the moving mode. This is probably the reason why in Experiment 2, where the pace is controlled by the user, this pace is significantly faster than in static mode. This increase in pace does affect the recall, and that is not strange since we know from Experiment 1 that an increase in pace causes a decrease of recall, irrespective of the mode. A fast self chosen pace will as a result also go along with a loss of recall. We can also confirm this with the average number of hits (true positives) per second. This metric appears to be roughly constant for both presentation modes in Experiment 2.

The preferences for the modes differ strongly, and quite some opposite arguments were used by the participants. It is therefore hard to make a clear statement on which mode maximizes the user satisfaction. We indirectly measured the perceived pace and confidence, which is in favor of the moving mode. We observe two equally sized groups: The first group likes the moving mode because it allows for an efficient vertical scan and a reduction of eye movement. The second group dislikes the moving mode because they have difficulties seeing the images because of motion blur or just feel nauseous because of the movement. So with respect to user satisfaction there is no clear winner, and we cannot confirm Hypothesis 2. There is however no doubt that the ability to control the pace increases the satisfaction, as this was strongly appreciated by all participants in the second experiment.

The fatigue effect cannot be measured clearly. We see a slight decrease of recall in the second part of Experiment 1, but this is not statistically significant if confidence bounds of 95% are used. In the longer sequences of Experiment 2, which contain 2400 images each, the user controlled pace is more or less stable. There are no signs that at the end of the sequence, participants structurally felt the need to reduce the pace. At eight images per second, which is a typically chosen pace in static mode, a sequence of 2400 takes five minutes to scan through. So for time intervals of a few minutes, we can reject Hypothesis 4. Below is a summary of our findings regarding the hypotheses:

- H1 The recognition accuracy is similar for both modes, under the assumption that the same pace is used.
- H2 The participants had too diverse preferences for the mode for a definite conclusion.
- H3 When controlling the pace, participants significantly worked faster in moving mode. The number of hits per second remained however roughly the same.
- H4 No significant signs of fatigue could be measured, especially not during the +/- 5 minute sessions in Experiment 2.

7.2 Limitations and future work

When images are moving at a high velocity they tend to look blurry, an effect that is caused by the monitor and is called *LCD Motion Blur*. We conducted the experiments on a laptop with a regular monitor that suffers from this effect. One could argue to do the experiments using a high end monitor with a low response time, but since LCD monitors are so ubiquitous, the results of our study are still relevant. Nevertheless, it would be interesting to see if a better monitor increases the usability of the moving mode.

There are two potential threats to validity we would like to mention here. The first one is that in the second experiment, moving mode always preceded static mode. The learning effect could potentially have influenced the results here. Furthermore, the data set is quite small in this study: 18 participants performed 12 sequences in Experiment 1 and 2 long sequences in Experiment 2. As a result, it is hard to analyze possibly interesting correlations between age, gender, etc. and the effectiveness, efficiency, and satisfaction. With our data, we could not measure such correlations. More (samples per) participant(s) could give better insights in future studies.

In all experiments, the precision was relatively high and that was probably due to the configuration of the image collection. By filtering the images from the noise collection that are closely related to the target categories, and selecting only clear target images, there were not many false positives. We even think that most of the false positives were caused by hitting the SPACE bar too late. This could affect several statistics like recall, response time, and the precision itself; but given that the precision is so high, this influence must be negligible. Another interesting aspect is the large difference in individual results, as is revealed by the large standard deviations in most of the graphs. One possible explanation is the limited number of targets in each sequence: missing one target image has a large impact on the recall for that sequence. We see similar effects in related work [1][10]. Adding more target images might make the analysis more robust.

Finally, we mention two more aspects that could be investigated more extensively in future experiments. First, we found that the preferred pace in moving mode does not decrease, even after a couple of minutes. It would be interesting to see how the pace develops in longer sequences, for example one that lasts one hour. Finally, we mentioned that the number of rows influences the required horizontal speed. A reduction of this speed can potentially take away the user's complaints about the moving mode. A reduction of the thumbnail size could hence be beneficial, but can also make it harder to recognize the images.

8 Conclusions

We have investigated the usability of two presentation modes for an image classification task, and the influence of pace control. Our conclusion is that the mode has no significant effect on the precision and recall when the pace is fixed. However, the moving mode is perceived slower and humans have more confidence when using this metaphor. This causes users in moving mode to choose a faster pace when they are enabled to. Since an increase in pace goes along with making more mistakes, users will as a result make more mistakes in moving mode than in static mode. The differences between the modes do not influence the average number of targets found per second.

For designers of visualizations that have to support the search of images or other visual objects, we give the advice to use static mode if high recall is necessary and to use moving mode for quickly scanning a large collection of visual data in order to get a rough overview. In case user comfort plays a critical role, for example if the task has to be done for long time intervals, then it seems a good idea to provide both modes and let the user choose. From their feedback we learned that users have clear but different preferences with regard to the mode. The performance of the moving mode can potentially be improved by using a monitor that does not suffer from motion blur.

Acknowledgments

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

REFERENCES

[1] T. Brinded, J. Mardell, M. Witkoski, and R. Spence. The effects of image speed and overlap on image recognition. In *International Conference on Information Visualization*, pages 3–11, July 2011.

[2] O. Christmann, N. Carbonell, and S. Richir. Visual search in dynamic 3d visualisations of unstructured picture collections. *Interacting with Computers*, 22:399–416, Feb. 2010.

[3] K. Cooper, O. de Bruijn, R. Spence, and M. Witkowski. A comparison of static and moving presentation modes for image collections. In *Proceedings of the working conference on Advanced Visual Interfaces*, May 2006.

[4] S. Corsato, M. Mosconi, and M. Porta. An eye tracking approach to image search activities using rsvp display techniques. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 416–420, May 2008.

[5] K. I. Forster. Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, 8:215–221, 1970.

[6] A. G. Hauptmann, W.-H. Lin, R. Yan, J. Yang, and M.-Y. Chen. Extreme video retrieval: Joint maximization of human and computer performance. In *ACM Multimedia*, pages 385–393, Oct. 2006.

[7] M. J. Huiskes and M. S. Lew. The MIR Flickr retrieval evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval*, pages 39–43, Oct. 2008.

[8] International Organization for Standardization. *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*, 1998.

[9] H. Luan, Y.-T. Zheng, M. Wang, and T.-S. Chua. Visiongo: Towards video retrieval with joint exploration of human and computer. *Information Sciences*, 181(19):4197–4213, May 2011.

[10] J. Mardell, M. Witkowski, and R. Spence. A comparison of image inspection modes for a visual search and rescue task. *Behaviour & Information Technology*, pages 1–14, Aug. 2013.

[11] J. Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers Inc., 1993.

[12] M. Porta. Browsing large collections of images through unconventional visualization techniques. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 440–444, May 2006.

[13] B. Spence, M. Witkowski, C. Fawcett, B. Craft, and O. de Bruijn. Image presentation in space and time: Errors, preferences and eye-gaze activity. In *Proceedings of the working conference on Advanced visual interfaces*, pages 141–149, May 2004.

[14] R. Spence. Rapid, serial and visual: a presentation technique with potential. *Information Visualization*, pages 13–19, 2002.

[15] R. Spence and M. Witkowski. *Rapid Serial Visual Presentation: Design for Cognition*. Springer London, 2013.