Quantity versus Quality – The Role of Layout and Interaction Complexity in Thumbnail-based Video Retrieval Interfaces

Wolfgang Hürst
Department of Information and Computing Sciences
Utrecht University, Princetonplein 5,
3584 CC Utrecht, The Netherlands
huerst@uu.nl

Dimitri Darzentas
Department of Information and Computing Sciences
Utrecht University, Princetonplein 5
3584 CC Utrecht, The Netherlands
d.p.darzentas@students.uu.nl

ABSTRACT
In this paper we present our findings from an experiment designed to test the effectiveness of complex, thumbnail based layouts and the interaction methods required, in video retrieval scenarios for handheld devices such as smartphones. Our evaluation explores the relationship between the number of thumbnails (Quantity) visible on screen and their discernible detail (Quality) with regards to the related necessary amount of interaction. The results indicate that such layouts achieve a very high rate of accuracy and speed but that the nature of the interaction is of critical importance for the success of the system.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Evaluation/methodology, graphical user interfaces (GUI), screen design, style guides, user centered design*

General Terms
Design, Experimentation, Human Factors, Theory

Keywords
Mobile Video, Video Retrieval, Thumbnails, Mobile Cognition

1. INTRODUCTION
Thumbnail images have become ubiquitous in our daily digital life as representations of, for example, larger photos or video clips. The old adage; "a picture is worth a thousand words" is well and truly verified in this regard. Ever since the thumbnail first replaced the icon as a visual preview representation of a digital image (or video) file, it has become accepted as the de-facto standard practice. The ability of humans to perceive and recognize features in these small pictures has made them an important tool in non-automated video browsing [10]. A simple small image can convey large amounts of information in an instant. As a fast and effective way of visually browsing (with no other assistance such as metadata) through image or video items, thumbnails are undisputedly one of the most effective methods [16] hence their widespread adoption in all forms of digital galleries, both on desktop systems and on mobile handheld platforms.

With the latter becoming the most ubiquitous connection with the digital world, especially in form of smartphones, the question arises if and how we can create efficient and effective video search and retrieval interfaces given the unique form factor and small screen sizes of such devices. On desktop systems, interfaces for video retrieval have been heavily reliant on the presence of metadata accompanying the video data, and the design of these interfaces reflects this [5]. Results from video queries are most often returned as a list of videos, presenting each with its relevant metadata and accompanied by a single descriptive thumbnail. Other approaches, such as the popular CrossBrowser implementation [15] focus on the presentation of multiple thumbnails extracted from individual videos and representing their respective content. Individual videos are often represented by a temporarily sorted sequence of extracted thumbnails – commonly referred to as storyboards [1, 6].

While usability and usefulness of such interface designs have been verified by many scientific studies in desktop system environments, their applicability for popular smartphones seems questionable due to their much smaller screens – typically around 3.5 to 4 inch diagonal versus 17 inch or larger diagonal on desktop systems – and different interaction modes – typically touch screen interaction versus mouse and keyboard on desktop PCs. Motivated by the relevance and importance of thumbnails for video retrieval interfaces, [7, 8] present a series of experiments addressing the first issue, i.e. what are optimal sizes and type (static or dynamic) of a thumbnail on a mobile platform? Their encouraging findings showed the use of thumbnails is actually a very effective means of data retrieval on mobile platforms, despite their limited screen size. However, related tests were limited to individual thumbnails shown in isolation [7] and horizontally presented strips [8]. It is yet unclear if and how they generalize to more complex layouts as they are required in realistic video retrieval tasks. In addition, the experiments presented in [8] suggest that the need for interaction might hinder users from utilizing thumbnails in the optimum way that was identified in [7]. Hence, using these findings as a starting point, we apply their conclusions to more elaborate and real-world scenarios to test whether the results can be transferred from single solitary thumbnails and small strips, to more complicated layouts, in order to determine the next best step for the effective design of mobile video retrieval interfaces. In particular we were interested in the effectiveness of instance recognition tasks [13] when attempted in combination with these layouts.

Given the high popularity and relevance of storyboards for video retrieval, we decided to focus on a grid (or matrix) layout of thumbnails in this paper (as illustrated in Figure 1). This layout is applied to a group of thumbnails extracted from a single video clip in a time ordered sequence arranged on a grid. This

ICMR'12, June 5-8, Hong Kong, China

conventional layout is also familiar from reading text left-to-right, bottom-to-top (at least in the case of western texts). Its use has also been endorsed by TRECVID [6, 13]. Focusing on search within a single video file was motivated by the related retrieval task, which will be discussed later.

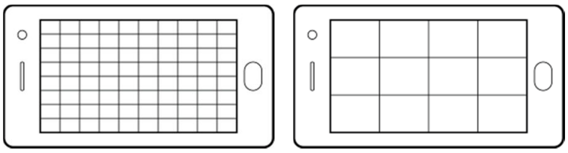

Figure 1. Storyboards-like grid-based thumbnail layouts

The parameters of a grid of thumbnails are essentially the size of the thumbnails themselves and immediately tied to that variable is their quantity. The smaller the size of the thumbnails, the more of them can fit on the grid. This leads to a ratio of size to quantity. What this means is that we can either have a lot of small images, and therefore greater granularity of information, or fewer larger images, which while more detailed individually, may not give enough information overview about the content of the video clip.

One way to break this stalemate is to lift the limit of the number of thumbnails by making the grid scrollable; this has the immediate benefit of allowing as many thumbnails as deemed necessary and therefore maintaining a finer granularity without necessitating smaller thumbnail sizes. The downside is that not all the thumbnails are displayed at the same time and an additional layer of physical and cognitive interaction is necessary. Figure 2 shows one possibility of such an arrangement and the one we chose to pursue. In our experiment we utilized "continuous scrolling" identical to the methods used by desktop applications instead of the "discrete scrolling" methods that enable a page-wise navigation and are often employed by mobile devices to, for example, switch between different "screens" of icons and digital book readers following the appropriate page metaphor.

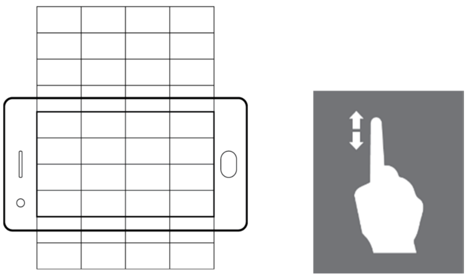

Figure 2. Scrollable grid layout for browsing larger quantities of thumbnails

This paper presents the findings of an experiment we set up to study the effects of the above variables (size, quantity, grid type) on the effectiveness of instance recognition tasks. We also evaluated the related interaction to determine whether such a setup would be appropriate and useful to real-world users and applications, such as video browsing for entertainment. By testing a 'bare-bones' static grid, and then systematically adding features and functionality, in this case the ability to scroll through more content, we plan to roadmap the creation of an effective thumbnail based layout for video retrieval on mobile handheld devices.

It should be noted that we were not immediately aiming to determine the optimal technical parameters for the thumbnails of a grid based on retrieval performance but to understand how well the layout facilitates effective video browsing and video retrieval tasks. In the following sections, we shall discuss the related work (section 2), elaborate further on the details and methods of the experiment (section 3), provide an analysis of the results (section 4), the conclusions, and related work (section 5).

2. RELATED WORK

Generally, we can assume that providing as much ancillary information about the content of a video as possible with a query result will lead to better search performances. This intuitive statement has resulted in many effective but rather complex interfaces which take full advantage of the large screen sizes (see [11] for a related overview). Considering video retrieval interfaces for contemporary mobile platforms, such as smartphones, which are fast becoming commonplace and are a good indication of current interaction technology, the situation is rather more complicated. Small screen sizes severely limit the expansiveness, complexity and therefore usefulness of advanced video retrieval devices. Indeed one may argue that these limitations could severely impact the effectiveness of the thumbnail representation as one of the most effective tools in the video retrieval arsenal.

Thumbnails have been accepted as the standard base on which nearly all video retrieval interfaces are built. Even the most barebones, text-query, interfaces strive to provide a thumbnail of the retrieved video file. Other thumbnail-based interfaces often use metaphors like filmstrips [3], Manga-like collages [2, 4], and the aforementioned storyboards [1] to represent the content of a video to the searcher. Thumbnails in all their forms are unquestionably effective on desktop systems for video retrieval. Motivated by the instinctive concerns about their size and practicality on small mobile screens [7] presented an evaluation where users were presented with different sizes of single solitary thumbnails in relation to common video retrieval tasks. The experiments also included the type of the thumbnails as a factor, examining whether static or dynamic thumbnails were more effective. The conclusions showed that rather small images can still be reliably used for search and target classification and if the thumbnails are dynamic even much smaller still lead to a reasonable retrieval performance. In particular, thumbnail sizes as small as 80 and 60 pixels for static and dynamic ones, respectively, have been determined as a viable conveyance of information on a mobile platform.

However, in a direct follow up experiment, [8] identified that these results are only valid in the case of a solitary thumbnail representation, but do not apply for designs in which single linear stripes of five to nine thumbnails are shown. In particular, the reported experiments suggest an optimum thumbnail size of 110 pixels, which still seems much smaller than originally anticipated. Maybe even more important, the studies revealed that users are not taking advantage of the option to play individual thumbs within the filmstrip as dynamic thumbnails – despite their superiority for search tasks that was demonstrated by [7]. Reasons for this might be a combination of a higher cognitive load resulting from the additional need to interact as well as the fact that more information about a video's content was shown at a time (thumbnails in the test have been extracted from a single video file).

Although providing encouraging information for the design of more advanced interfaces for video retrieval via mobile platforms, the results presented in [7, 8] leave several important questions unanswered. In particular: If optimum thumbnail sizes increase when we switch from classifying solitary thumbnails to film strip representations, will storyboard-like matrix representations such as depicted in Figure 1 result in a need for even larger thumbnails? And if so, how does the resulting need for interaction (cf. Fig. 2) influence retrieval performance and subjective search experience – especially considering the negative experience with interaction reported in [8]?

In order to further investigate the effects that more complex layouts and arrangements would have on the users and their ability to complete certain video browsing tasks, we present an experiment evaluating the parameters thumbnail size in a grid-based layouts as shown in Figure 1 under consideration of potential issues resulting from scrolling interactions as illustrated in Figure 2. We define interaction here as both retrieval performance and as how well the interface facilitates effective browsing [5]. For our tests, we decided to restrict them to static thumbnail representations, based on the same arguments as [8], i.e. that simultaneous playback of various images as dynamic thumbnails would further distract users and most likely have an impact on performance and overall effectiveness when searching for information. The major parameters to explore in our test are thumbnail size and the related ability for users to find content represented by them, and the role interaction has on the effectiveness of the interface. Because these are independent of the actual source of the data represented, we only tested search in single video files, but expect our results to generalize to search in multiple videos, where, for example, each row in the matrix represents the time-ordered sequence of thumbnails extracted from single videos.

3. EXPERIMENT

3.1 Interface Rationale & Design

To begin investigating the effects of presenting multiple thumbnails simultaneously on video browsing, a number of different layouts were considered, including grids, linear scrollable strips, three dimensional perspective representations and other less conventional arrangements. As a first step we settled on a grid layout reminiscent of a traditional storyboard, a layout that is intuitive and effectively communicates it purpose. It caters to user expectations as it follows the conventional left to right and line by line reading styles prevalent in all (western) media, such as printed works, comics, etc.

More importantly, the grid layout also allows for the maximum quantity of thumbnails possible in a two dimensional interface without any overlap. Specifically, for the static grids, the number of possible thumbnails that would fit on the screen of the device was determined by the dimensions of the thumbnails. For the smallest size thumbnails (80 pixels) the number of thumbnails was 10x8 = 80 images. This amount decreased progressively as the thumbnail sizes increased. For the largest (200 pixels) images, only 12 images could fit on-screen. The illustrations shown in Figure 1 illustrate the exactly the largest (left) and lowest (right) number of thumbnails used in the experiment.

In the case of the scrolling grid (cf. Figure 2), while there was no upper limit on the total amount of thumbnails, the quantity that could be displayed on screen at any given time was still held by the same restrictions. For consistency, across all thumbnail sizes we limited the total amount of thumbnails so that the total "scrolling distance" would be about two and a half single screen grids.

3.2 Experiment Design

In order to determine the effectiveness of the grid based layout, we adapted the experimentation methodology of [7] and set up an experiment to progressively test the ability of the user to complete instance detection tasks. Grid layouts were evaluated by the subjects in decreasing number of thumbnails (i.e. increasing thumbnail sizes) in order to benefit from learning effects and better identify subjective user perception of prefered thumbnail sizes (cf. [7]). A series of questions/tasks would be asked of the user with the parameters of the layout changing every few questions.

The layout parameters under examination are the size of the thumbnails, their number and whether the grid is static (limited to the current screen) or scrollable (bigger than the current screen and therefore scrolling is necessary in order to view all the thumbnails). An emergent parameter that is under scrutiny is the ratio of the thumbnail size to the thumbnail quantity. Determining the ratio that corresponds to the optimum balance between information depth and information breadth is one of the key issues in a related interface design (quantity – quality).

The profiles of the users are of course also an important factor. Therefore the experiment was designed to gather and catalogue the information from the users both prior to and after the experiment was conducted. Prior to the test, the users were queried on typical relevant factors such as gender and age and also on their attitude and experience towards technology, mobile devices and the use of technology for entertainment. After the experiment, the users were queried again, this time on their opinion of the interaction. Importantly they were specifically not asked for their opinion on the thumbnail's parameters, e.g. "which size did you find more convenient?" because it seems obvious that users would most likely agree that larger thumbnail sizes are better suited to identify the contained content. Instead, they were asked questions regarding the interaction method itself, e.g. "How intuitive/natural did you find the interaction?" because the need for more interaction activity is a natural trade-off between thumbnail size and the resulting number displayed at a time.

3.3 Implementation

To perform the experiment we settled on the creation of a single mobile application on the Android platform. We implemented a portable mobile application that would serve as a distributable test to gather data. The application encompassed the entirety of the experiment, including questionnaires, feedback mechanisms and the performable tasks themselves. It was designed in such a way so that it would serve as a modular, expandable and flexible test system for the possibility of future adapted tests or different experiments. It was also designed to be autonomous, gathering and dispatching the data to the researchers automatically to provide the additional option large scale distribution for gathering with large numbers of users should the need arise. Figures 3, 4 and 5 show the end result of the implementation. Thumbnails in the figures have been replaced for this publication with non-copyrighted material, but reflect the type and style of data that was used in the tests.

The source video data that was used to extract the thumbnails was a collection of various episodes from contemporary television series. The wide variety and today's prevalence of television media provided a consistent and convenient base of data that directly relates to real world use cases. The thumbnails for each task were extracted from different video clips. The choice of which thumbnails to extract from each video was based on the requirement that they should cover the whole length of the clip and be equidistant in frames.

Figure 3. Screenshot of the layout used in the experiment (with largest number of thumbnails)

Figure 4. Photo of the layout used in the experiment (with the smallest number of thumbnails)

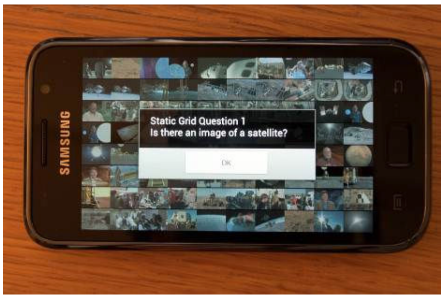

Figure 5. Photo of the experiment procedure (with a question being asked via a text label popup)

The tests that we report on in this paper were all administered and supervised. This was done so that in addition to the quantitative data gathered by the application, we could also obtain some qualitative data. The quantitative data was in the form of transaction form analysis and task time, while the qualitative data arose from the observed user reactions and voluntary think-aloud opinions from the participants.

The overall testing process begins with a short questionnaire establishing details such as age, gender and experience with mobile technology. This is followed by a short onscreen tutorial that familiarizes the users with the imminent testing process by running them through some non-interactive mock questions with popups explaining the functionality and options. The users are then presented with a sequence of 24 questions. Each question is an instance recognition task (cf. below). The users peruse the thumbnails and can highlight any number of them that they consider as applicable. The users may also provide a negative answer when they are sure that the requested instance is not contained in the video. In the case where the users feel that they cannot commit to an answer, i.e. they are unsure, there is a third option available to them – i.e. a button indicating "unsure".

Tasks were motivated and designed in accordance to the 2011 TRECVID "Instance Search" category. Examples for typical search motivations in relation to the used TV shows include situations such as: "I want to re-watch that funny scene from the show that I saw some time ago, where the one character was wearing this funny hat," or "Is this the episode in which this character was wearing that funny hat?" However, since the participants were unfamiliar with the actual content of the data used in the tests, we phrased the questions neutrally. For practical reasons, they were also formulated as yes/no questions that could be answered by "I found it/It doesn't exist/ I cant tell", resulting in tasks such as "Is there a man wearing a read baseball hat in the video?" or "Is there a fire truck in the video?"

Of the 24 questions (or tasks), the first 12 are on static thumbnail grids and the rest are on scrollable thumbnail grids. The size of the thumbnails begins at 80 pixels, equivalent to 8.7mm in width, and increases in steps every 3 questions, finally reaching a size of 200 pixels, equivalent to 21.5mm. In the case of static grids this means that a single screen on the first 3 tests contains 80 thumbnails, this number finally shrinks to 12 in the last three tests. In the case of the scrollable grids, regardless of thumbnails size, the user is able to scroll downwards approximately two and a half screens.

The sequence of tasks was the same for all the participants so as to clearly indicate any trends and inconsistencies. We chose to have the progression of the tasks (small to large thumbnails) the same for all the test subjects, so that any learning curve can be identified and taken into account. Upon answering the final question, the user is presented with a closing questionnaire that queries them solely on the interaction experience.

4. DATA ANALYSIS AND RESULTS

The data collected from the supervised tests, when collated and analysed, presented very encouraging and interesting results. As the main distinction between the tasks was the type of grid (Static or Scrollable) it is beneficial for ease of comprehension to separately assess the results for each case.

Before moving on to this separate analysis, a few observations must be made on the users. We chose to collect a sample that was as balanced as possible, gender wise, with 14 male and 10 female participants. Tables 1 and 2 show their answers to questions regarding their experience with technology and more specifically mobile platforms.

Table 1. Relationship with Technology

Positive - Everyday use	8
Neutral - Circumstantial use	11
Negative - Avoided if possible	5

Table 2. Relationship with mobile platforms (smartphones)

Positive - Part of everyday life	11
Neutral - Dictated by circumstances	4
Negative - Avoided if possible	9

The tables illustrate the fact that many of the participants did not consider themselves experienced users, what would be colloquially known as “power users” and that some of them even have negative relationships with technology. What is of great interest though is that all 24 users, when asked the question about whether they used mobile technology for entertainment, answered without any observed hesitation, yes. This observation also holds across all age groups as the sample that was tested was distributed over a large range of ages, as can be seen in Table 3.

Table 3. Numbers of participants per age group (top row)

21-30	31-40	41-50	51-60
8	9	4	3

As previously mentioned the use of contemporary technology is not limited to a small “technological elite” but is a phenomenon widespread not only across multiple knowledge levels but also across age groups and generations of users. Indeed the fastest growing demographic for the use mobile technology is the baby boomer generation (55-64 year olds) [14]. Therefore we purposefully sought out users that would belong to the more advanced age groups, who have had to deal with the advances of technology over the course of their lives in favour of younger users who have become acclimatized to today’s technological metaphors.

In the following sections, we shall have a detailed look at the individual results for the static and the scrolling grids and their comparison. Each question asked of the users is referred to as a task, with 24 tasks in all, 12 for the static layout and 12 for the scrolling layout. Each set of 12 is divided into 4 groups which share the same size parameters. Finally we analyse the user feedback on the interaction experience.

4.1 Static grid

For the case of the static grid it was apparent early on in the analysis that the levels of accuracy were high and very consistent. Indeed the average rate of success across all 12 tasks is 85% with half the users scoring above 87.5%. Within the static tasks the main variable parameter is the size and quantity of thumbnails. With the first 3 of the 12 static tasks featuring very small (80 pixels wide) thumbnails, the success rate began at a moderate 66% for the first task and 70% for the second, rising immediately for the third one as illustrated in Figure 6, which also shows the results for the three tasks of each other thumbnail size (100px, 133px, 200px, respectively).

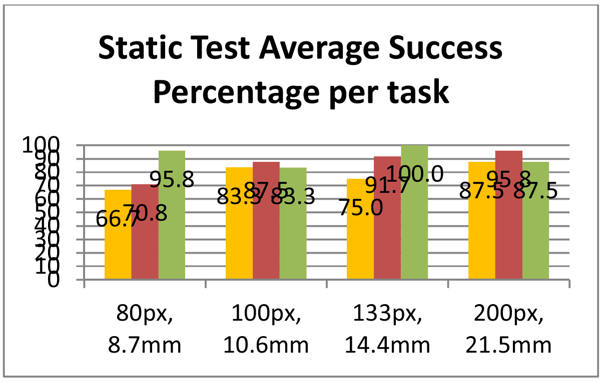

Figure 6. Success rates (static test, detailed results)

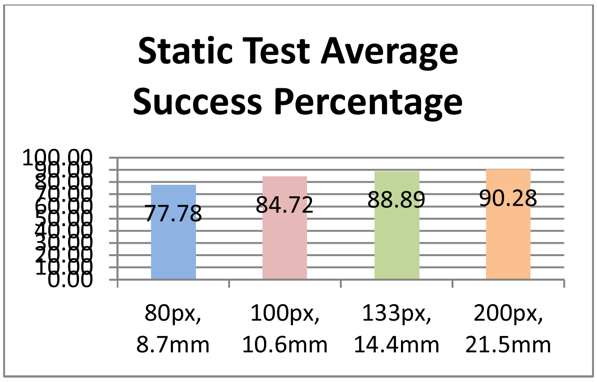

Figure 7. Success rates (static test, averages)

Figure 7 illustrates how, despite the lower performance of the first task start, the success rates rise fast, and continue rising throughout. Each bar represents the average success rate for each thumbnail size task group, 3 tasks for each group for a total of 12 static grid tasks shown in Figure 6. A plausible explanation, the low score on the first task is a necessary acclimatisation period by the users. As mentioned above in section 3.3 we anticipated a learning curve.

Obviously, the improvement in performance was to be expected as larger thumbnail sizes should result in a better visibility and thus better classification performance. It is surprising though that in contrast to the relatively large increase of about 7% from 80px to 100px thumbnail sizes, further improvements are less dramatic, i.e. only about 4% and 2% from 100px to 133px to 200px. This is particularly noteworthy with respect to the results presented in [7, 8] which show a strong increase in optimum thumbnail size when moving from solitary thumbnail representations to film strips. Apparently, this does not seem to be the case when moving from strips to full, matrix-like storyboards.

However, besides the high and consistent level of accuracy, also of some note was the drastic reduction in time required by the users to make a choice as the size of the thumbnails increased. Figure 8 illustrates this result with each bar again representing a thumbnail size task group. The overall average time to complete a task across all the static questions was 25 seconds.

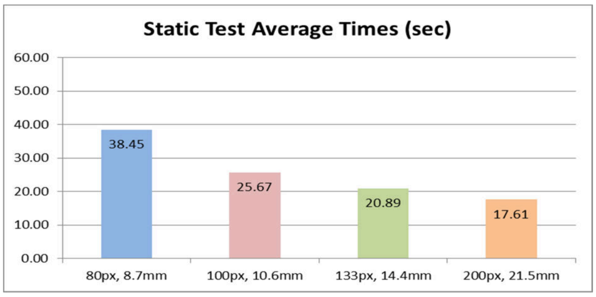

Figure 8. Time to solve the tasks (static test, averages)

The decrease in the time to answer might have been expected as the increasing thumbnail sizes naturally lead to smaller quantities (less individual thumbnails to look at) of thumbnails. However as the thumbnail sizes increase, so does the amount of visual detail that must be processed (bigger, more detailed images). It would appear that it is faster and easier to analyse fewer, more detailed images, than it is to scan through several less detailed ones.

As the experiment was designed so that some of the questions would have a number of correct answers, there are occurrences where some users did not detect all the correct answers, therefore the success rating for each individual task is measured as a percentage with a 100% score representing the case where a user has found all the applicable instances. Figure 9 illustrates the comparison of those who got at least one right answer and those who got all the right answers for each given task.

As is apparent from the graph, for the most part, users consistently found all the correct answers for each task, with a single outstanding outlier, which has been attributed to the relative ambiguity of the question.

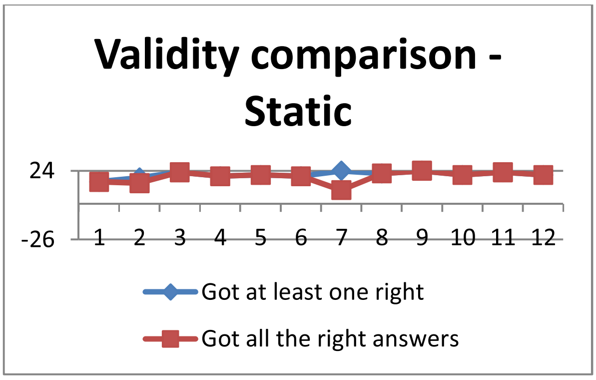

Figure 9. Correctness of the results (static test)

Therefore in the case of the static grid, we can state that the layout appears to be an effective, fast and accurate method of expediting instance recognition tasks. For real world applications though, these findings are not directly applicable, as the limited numbers of thumbnails that can be concurrently displayed on screen at any given time are too few for extensive video search tasks. Therefore, we move on to the next part of the experiment, lifting the limit on the quantity of thumbnails available (but not displayed) by using a scrollable grid.

4.2 Scrolling grid

The results of the scrolling grid layout tasks paint a different picture than for the static cases. The average success rate, across all the thumbnail sizes remained good, but lower than the static grid results, with a final average of 61% and half the users scoring below 56%. Figure 10 shows the success rates on the individual tasks for the scrolling grid.

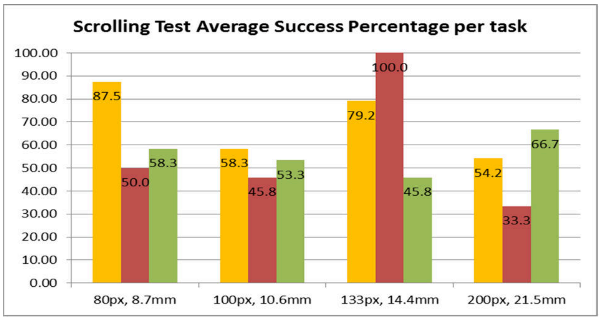

Figure 30. Success rates (scrolling test, detailed results)

Figure 11 below illustrates the average success rates for the 4 groups of thumbnail sizes. In the case of the third group (133 pixels) an outstanding outlier occurred where all users scored 100% on one of the given tasks, thus explaining the comparatively high success rate for thumbnail sizes of 133px. Although the otherwise observable decrease in performance for larger thumbnail sizes seems low, a comparison with the related increase of performance in a static scenario (cf. Fig. 6 and 7), confirms our expectation that the need for interaction can have a negative effect on retrieval performance.

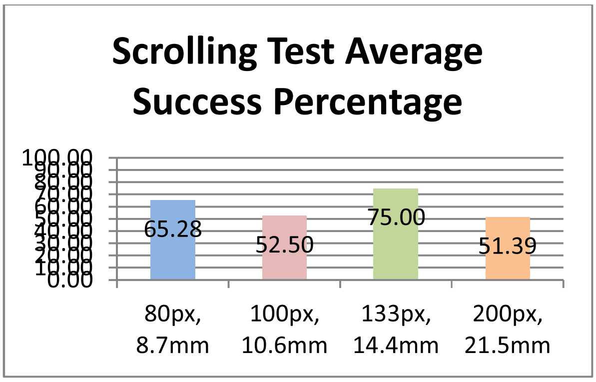

Figure 41. Success rates (scrolling test, averages)

The time performance was also lower in comparison with the static grid tasks, with the average time required to complete each task rising to 34 seconds. Figure 12 shows the average time required per task according to thumbnails size. The higher times can be attributed in some degree to the fact that there were physically more thumbnails for the user to peruse, approximately 2.5 times more. If the physical amount of time spent interacting with the interface, i.e. 'swiping' is taken into account and not added to the total time taken to answer then time difference between the static and scrolling layout tasks is diminished.

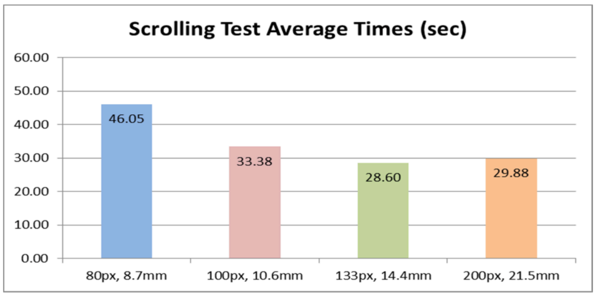

Figure 52. Time to solve the tasks (scrolling test, averages)

As far as the validity comparison results for the scrolling grid tasks, Figure 13 illustrates that, unlike in the static grid tasks, the users who did actually succeed in correctly completing the tasks found all the right answers for those tasks and not just some of them (hence there is no blue line on the graph). However the task success rates remain lower than for the static cases. Therefore, we can speculate that uncontrolled, continuous, non-discrete, scrolling is not an immediate 'upgrade' to the static layout and that further refinements must be made to maintain high success rates and lift the restrictions of the static layout.

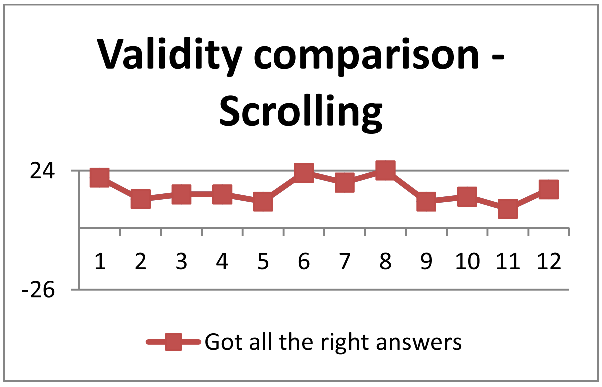

Figure 63. Correctness of the results (scrolling test)

4.3 Post-test user feedback

Immediately after finishing the 24 tasks set to them, the users were presented with a short questionnaire querying them on the experience and opinion of the interaction. There were no direct questions regarding the parameters of the tasks, such as eliciting preferences for thumbnail size and quantity, as the aim of this questionnaire was to determine how whether the setup was 'usable' from a human centric point of view. The following Table 4 illustrate their answers.

Table 4 Qualitative user feedback.

	Negative	Neutral	Positive
Was the interaction intuitive?	4	9	11
Was the interaction enjoyable?	3	14	7
Would you use it?	7	9	8

On the whole, the results show that the users' opinions did not indicate any significantly strong trends. There were not many downright negative opinions regarding the intuitiveness of the interactivity, but a noticeable amount of users rated interaction lower with respect to fun and enjoyment. Interestingly however, most of them also expressed their dislike about the images being too small in the settings where less interaction was required. Several users also expressed the opinion, during and after the test, that they found the static grid more comfortable and intuitive. Indeed it was apparent to the observer that nearly all the users preferred to tackle the problem as an isolated onscreen challenge.

Despite the more neutral rating regarding the characterisation of the interaction, a majority of users declared that they would use such an interface. Overall, user opinions indicate that a grid layout could indeed be a welcome part of video retrieval interfaces; this is also well substantiated by the quantitative data; despite the expressed misgivings of some of the users, their success rates were consistently high, as seen above in the data analysis. Even in the scrolling grids the success rate rarely dropped below 50%.

Although many users, especially those belonging to the advanced age groups, expressed that they felt daunted by the tasks that utilized the smallest size thumbnails and worried about the accuracy of their scores. However, the data showed that they fared no worse than users from younger age groups. We expected some variance of the success rates due to age related factors (eyesight, experience, etc.) but these were not immediately apparent in our test sample and will be revisited in the future with larger samples.

5. CONCLUSIONS & FURTHER WORK

Our analysis shows that the static thumbnail grid layout achieved and maintained very high accuracy over all the different thumbnail sizes. Compared to previous works in this area [7, 8], it is clear then that the effectiveness of thumbnail based interfaces that was observed in the related tests is successfully, if not quite fully, transferred to more complex layouts. Most importantly, a similar increase of optimum thumbnail size from 80 to 110px as reported by [8] when moving from solitary thumbnails to film strip representations could not be observed when moving to the even more complex matrix-style storyboard layout – where thumbnail sizes of about 130px achieved an equally high performance rate of 90% as the much larger 200px size version of the interface. This is particularly encouraging, since the layouts evaluated in our experiment can be far more applicable and useful to real world scenarios.

Unfortunately, these same levels of successful results were not similarly observed in the case of scrolling grid layouts, where the added layer of interaction complexity took its toll and rendered the system less effective. This confirms our assumption about the critical influence of the interaction design on retrieval performance. It is in line with the observation reported by [8] that users did not interact as much with the system as earlier results reported in [7] would suggest. The difficulties users experience with the scrolling interface can be likened to the issues that are apparent in all static peephole interfaces [9], where the additional spatial distortion of the content is added to the temporal distortion, further increasing the cognitive load of the user.

Therefore, it can be concluded that uncontrolled, continuous scrolling grid layouts are not a clear improvement over the static grid layout. Further refinements must be made to maintain the high success rates of the static layout and at the same time, remove its inherent restrictions. Possible options for further investigations include, for example, interface designs where users can switch between various static grid arrangements (similarly to swiping through different screens on modern smartphone interface designs) in contrast to the continuous scrolling evaluated in our experiment.

These conclusions of our study provide further insight into better use of complex thumbnail based layouts for mobile video retrieval tasks, but obviously, they are just one more step towards the ultimate goal of creating the perfect video retrieval interface for mobile devices. As already said above, interaction design is a crucial issue and the better, more intuitive, yet efficient and effective interaction modes need to found. Considering the actual layout, our results suggest that despite the small screen sizes of mobiles, even more complex interface designs such as scrollable linear strips, multi-tiered grids, and even 3D interfaces [12] could be possible. Likewise, the usefulness of dynamic thumbnails identified in [7] needs to be evaluated in more complex designs.

Further investigation is warranted into the exact specifications of the optimal balance between larger thumbnails (featuring more detail) and larger numbers of smaller (less detailed) thumbnails that a user has to analyze to complete a task. Is it faster and easier to examine fewer, more detailed images, than it is to scan through several less detailed ones or the opposite? The answer to this question is certainly dependent on the task at hand – indicating the need for further research about other retrieval than the ones studied here. For example, how effective are the proposed layouts for "event recognition" tasks that will gauge the ability of users to accurately detect events and series of actions in a video instead of just instances of entities. And how should different domains and contexts be considered in the interface design?

6. ACKNOWLEDGMENTS

This work was partially supported by a Yahoo! Faculty Research Grant. We would like to thank our test subjects for their time and willingness to participate in the experiment.

7. REFERENCES

[1] Adcock, J., Cooper, M., Girgensohn, A., & Wilcox, L. (2005). Interactive Video Search Using Multilevel Indexing. *In Proc. CIVR*, (pp. 205-214).

[2] Boreczky, J., Girgensohn, A., Golovchinsky, G., & Uchihashi, S. (2000). An Interactive Comic Book Presentation for Exploring Video. *Proc. ACM CHI*, (pp. 185-192).

[3] Christel, M.G., Warmack, A., Hauptmann, A.G., & Crosby, S. (1999). Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library. *Proc. IEEE Advances in Digital Libraries Conference*, (pp. 98-104).

[4] Christel, M.G., Hauptmann, A.G., Wactlar, H.D., & Ng, T.D. (2002). Collages as Dynamic Summaries for News Video. *Proc. ACM Multimedia*, (pp. 561-569).

[5] Christel, G. C. (2006). Evaluation and User Studies with Respect to Video Summarization and Browsing. *Multimedia Content Analysis, Management and Retrieval* 2006, Proc. of IS&T/SPIE Symposium on Electronic Imaging.

[6] Christel, M. G., & Yan, R. (2007). Merging storyboard strategies and automatic retrieval for improving interactive video search. *CIVR '07 Proc. of the 6th ACM International Conference on Image and video retrieval* (pp. 486-493).

[7] Hürst, W., Snoek, C. G., Spoel, W.-J., & Tomin, M. (2010). Keep moving!: revisiting thumbnails for mobile video retrieval. *MM '10 Proc. of the International Conference on Multimedia* (pp. 963-966).

[8] Hürst, W., Snoek, C. G., Spoel, W.-J., & Tomin, M. (2011). Size matters! How thumbnail number, size, and motion influence mobile video retrieval. *MMM'11 Proc. of the 17th International Conference on Advances in Multimedia Modeling* (pp. 230-240).

[9] Mehra, S., Werkhoven, P., & Worring, M. (2006). Navigating on handheld displays: Dynamic versus static peephole navigation. *ACM Trans. Comput.-Hum. Interact.* 13(4), (pp. 448-457).

[10] Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, (pp. 520-527).

[11] Schoeffmann, K., Hopfgartner, F., Marques, O., Boeszoermenyi, L., & Jose, J.M. (2010). Video browsing interfaces and applications: a review. *SPIE Reviews*, 1(1), (pp. 1-35).

[12] Schoeffmann, K., Ahlström, D., & Böszörmenyi, L. (2012). Video Browsing with a 3D Thumbnail Ring Arranged by Color Similarity. *Advances in Multimedia Modeling*, Springer, Berlin, Heidelberg, New York, (pp. 639-641).

[13] Smeaton, A. F., Over, P., & Kraaij, W. (2006). Evaluation campaigns and TRECVid. *Proc. of the 8th ACM International Workshop on Multimedia Information Retrieval* (pp. 321-330).

[14] Smith, N. (2011/08/23), Baby Boomers Up-to-Date on Smartphone Technology, Study Finds. *BusinessNewsDaily*. http://www.businessnewsdaily.com/1380-baby-boomers-smartphone-savvy.html

[15] Snoek, C.G.M., Worring, M., Koelma, D.C., & Smeulders, A.W.M. (2007). A Learned Lexicon-Driven Paradigm for Interactive Video Retrieval. *IEEE Transactions on Multimedia* 9(2), (pp. 280-292).

[16] Torralba, A., Fergus, R., & Freeman, W. T. (2008). 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (pp. 1958-1970).