Personalized Video Thumbnails

Matthew Fong, Kai Wu, Xueqin Zhang
University of British Columbia
Vancouver, BC, Canada
{mfong, kaywu, xueqin}@ece.ubc.ca

ABSTRACT

As more and more videos are being consumed over the web, organizing and searching for videos becomes more difficult. Videos in at-a-glance search spaces are generally represented by a title, a description, and static thumbnail to show the contents of the video. In the particular case where a user is looking for a video they have seen before, this information may not be enough to trigger a user's memory of the video. We introduce two techniques to represent video in a personalized way that is unique to the user through the analysis of their viewing behaviour, as well as explicit bookmarking for representative thumbnail selection. We performed a formal evaluation on the techniques compared with the traditional YouTube style representation and found users were faster in selecting the video using our methods, and also preferred our method.

Author Keywords

tutorial video; video search; video bookmarking;

INTRODUCTION

The dramatic increase in the quantity of videos people watch everyday, either in the online form of services such as Youtube, Vimeo or the more personal form of home video, causes challenges for individual users to browse their previously watched video list and find the specific ones they have watched. In general, video search is a complex and time-consuming task because people use mainly visual cues while current search engines mainly employ keywords to help users find a specific video, which is prone to failure due to mismatch with the video's metadata and not utilizing the rich visual information that is available in the videos. There is a lack of effective visualization of the video content that aids with the difficulty of searching and seeking information within videos. Currently, video thumbnails seem to be a popular method for showing video content to speed the searching and browsing. There are three major kinds of thumbnails seen in the field and in literature, which are static, slideshow, and interactive thumbnails [8]. In the interface of the search results, static and slideshow thumbnail are used to represent the video content, which is not necessarily the best representation of the video and most of the videos' visual information is lost. This is especially true for videos with no plot or story (home-made videos, lectures, tutorials), or with similar backgrounds and scenes (television episodes, lectures, etc). In this aspect, the interactive thumbnails are good solutions for representing videos as they let users explore the video as they may have remembered it, and previous research [3] showed that interactive dynamic video thumbnails are a better alternative for it provides more visual information with minimum extra work on the user' s part.

There are already many video browsing interfaces with interactive thumbnails currently. These thumbnails generally represent the entire video and allow the user to explore the entire video through mouse-over movement. This is a viable solution if it is a short video or there is distinctive/distinguishable content, otherwise it will take more time than if the thumbnails contain only what the user deemed to be interesting. Therefore, it is reasonable to come up with ways to personalize the thumbnails so that they display the desired content for each individual viewer.

There are generally two approaches to personalize the thumbnails so that they match the user's mental images and thus help to trigger the user's recollection of the specific video: 1) leverage each individual viewer's watching behavior and pattern to adaptively personalize the thumbnails; 2). utilize the crowdsourcing data uploaded by viewers with similar interests to personalize thumbnails for each video. A study in Chapter 3 of [2] showed that people have a tendency to re-watch video intervals. This user behaviour inspired us to use user's previous watching behaviour to personalize current interactive thumbnails to facilitate the user's video searching task. In order to utilize this typical user behaviour, we focus on videos where users are more likely to want to go back and rewatch certain segments and search/retrieve them after a certain period of time, like tutorial videos or lecture videos. Other work by Syeda-Mahmood and Ponceleon [9] have shown automatic thumbnail generation by crowdsourcing the "distribution of visit counts per unit time". This enabled thumbnails to be created that were more likely to be interesting to individual users, assuming that each user has similar interests. Personalizing the thumbnail generation to the individual user should, by extension, help the user find an interesting video they have previously seen.

In this project, we designed several interfaces that allow users to personalize the thumbnails of a video through manual operation and the behaviour they exhibit while watching the video.

Controlled experiments were conducted to verify the effectiveness of the thumbnail personalization methods. The main contribution is the idea of thumbnail personalization based on viewer's rewatching pattern.

RELATED WORK

We reviewed three bodies of work: *video representation*, which concerns different ways to visualize video content, *user-specific personalization*, which allows to cater to specific user through behavior tracking, and *revisitation pattern of web content*, which leverages the revisitation curves to predict and recommend web content for individual user.

Video Representation

Current solution of video visualization uses one static or dynamic thumbnail - images extracted from videos - to represent and summarize the videos. These thumbnails are organized in a linear/grid layout with static frames or dynamic animations. This approach provides a uniform interface for all users with no adaptation catering for individual difference, and suffers from lack of information at first glance. In this sub-topic, we focus on novel techniques that are designed to visualize video content.

Girgensohn *et al.* [3] took the interactive thumbnail and created the skimming interface with keyframe slider. The thumbnail shows one keyframe for each video and provides a slider that lets the user change the keyframes dynamically. The user changes the keyframe in the keyframe window by moving the mouse along a timeline. The keyframe closest to the cursor is displayed. A positive aspect to this approach is that it supports quick skimming to provide an overall impression of the videos context and temporal structure. However, only one frame is visible at a time which doesnt provide an overview of the videos content to the users at a glance. Furthermore, exploring a long video using this method can be tedious and cumbersome.

To better represent a video, Girgensohn *et al.* [3] extended their system with the Pictorial summary which uses keyframe data to determine importance. It calculates an importance score for each segment and threshold the score to prune less important segments. Keyframes are extracted for each selected segment and the sizes are emphasized or deemphasized based on the importance measure. If coordinated captions exist, they can be included in the display. Positive aspects to this approach are that it can show the content of the video in a concise and compact way, the keyframes and the timeline are synchronized so users can access a certain location using both approaches, and users can see both the visual representation via keyframe of an interested segment and the corresponding duration of the segment. The usage of automatic methods to find important keyframes is unreliable because it does not reflect what the user found interesting in the video. As such, there is no reflection of the video's importance from user feedback, and thus does not help the user in finding video they have seen.

Jackson *et al.* [5] used precisely animated arrangement slides thumbnails in their video surrogate system Panopticon, which displayed multiple sub-sequences in parallel to present a rapid overview of the entire sequence to the user. The video surrogate system achieved the overview of video through the animation of the timeline which displays all frames within an approximate 10 second time window. The interface of Panopticon is shown in Figure 4. Actually after their first user study, the authors find separating the dynamic timeline and video player wasted more time, so in Panopticon2, they integrated the video player into the dynamic timeline. A positive aspect of the system is that it can seamlessly search and seek information within videos, which is especially useful for surveillance videos and unedited videos. But in our general videos, especially ones we have watched, we just need to fast browse some specific video intervals to remind us of the video content.

User-specific Personalization

We are interested in two different kinds of personalization: user-driven personalization (adaptable) and system-driven personalization (adaptive). The former approach gives the user proper control over the system while there is cost in terms of time and resource and requires relevant skill sets. The latter approach frees the user from laborious and tedious work whereas may cause confusion and a sense of losing control. We focus mainly on adaptive personalization, and specifically, its application on automatic video thumbnails generation.

Greenberg's work on adaptive personalized interfaces claimed that users tend to retrieve items that have been accessed previously; they exhibit a repetitive pattern of access [4]. When it comes to retrieve a certain previously viewed video or video segment, users also need a personalized interface to help navigate through a large database more efficiently. Traditional use of thumbnails doesnt incorporate the users viewing behavior or pattern thus fail to present such a personalized interface.

In the space of personalized search, there has been work done to automatically identify users to cater results [7, 6]. This work has been applied to internet search, and is useful in behaviour tracking and allowing prediction on a per-user basis, and allows search results to be targeted toward the user based on all the behaviour listed in the work by Adar [1]. These methods can be extended to video search, with a different set of behaviours used as a criteria determining interesting video.

Revisitation Pattern of Web Content

One of the behavior that has been used frequently in the work of adaptive personalization is revisitation behavior. It gives insight to the user's intent behind this particular behavior and topics of interest, which can be utilized to trigger user's relevant recollection of the content.

Work by Adar *et al.* [1] has studied revisitation patterns in websites and made suggestions to the design implications they have found. By logging the inter-visit time which certain websites are accessed as well as the frequency, web browsers could suggest or predict URLs for the user. They also suggest search engine developers to preserve previously viewed results from repeat searches, or organize results based on how long a website was visited. These methods are excellent in

aiding users search through a personalized experience. The paper uses the revisitation curve generated from the inter-visit time to characterize the general pattern of revisitation of a specific set of Web pages. By grouping similar revisitation curves together, four visit patterns are discovered with associated sets of Web pages. These patterns are used to explain users' revisitation behavior and suggestions of improvement.

DESIGN RATIONALE

Here, we describe several design guidelines and methods that we have, through our experiments and interviews with users, employed into our prototype interfaces.

Representation of Thumbnails

Thumbnails are used to represent the content of the video, and so, they should be informative and customizable by each individual viewer to trigger relevant memory. As seen in [8], static thumbnails are not a good representation of the video content, thus we focus on interactive/dynamic thumbnails.

The thumbnail is the basic component of video representation, therefore they must be straightforward and intuitive enough to understand and interact with. In our first design sketch in Figure 8, the highlighted strips on top of the thumbnails represent the positions of the most watched video intervals within the video segment that this thumbnail represents. The second design sketch uses the same highlighted segments but only shows the frames within the highlighted video intervals instead of the whole video that this thumbnail represents. The idea behind the second design is that the most watched video intervals are more likely to trigger the viewer's memory, but due to the complexity and confusion reported by some users, we chose the first design sketch as our thumbnail representation.

Generation of Thumbnails

Since we want the thumbnails to be viewer-specific, we came up with three different ways to personalize thumbnails to the specific viewer: an automatic method that selects thumbnails based on viewer's viewing behaviour; a manual method that allows the viewer to manually select thumbnails from the video; and a mixed method that gives the viewer the option to manually change the thumbnails that are automatically generated from the most watched video intervals.

Automatic

Based on viewers' rewatching behavior, we can generate the most watched video intervals, from which the thumbnails will be extracted. The success of this method depends largely on the representation of the automatically generated thumbnails.

Manual

One of the flaws of automatically selecting thumbnails is the lack of proper representation of a video, which can be achieved through manual selection. The viewer can manually select any frame from the video during video watching. The drawback of this method is the intrusiveness and distraction, which interrupts the learning process and video watching experience.

Mixed: Automatic + Manual

The automatic method provides automation but lacks the ability to customize; the manual method guarantees customizability but is distracting and needs extra work on viewer's part. We came up with a design that combines these two, which makes a perfect balance between automation and manual operation with minimal intrusion and distraction.

PROTOTYPES

Throughout the development phase of the interfaces, we cycled through several iterations in the design of the interface elements. This section briefly outlines some of these designs.

Low-Fidelity Prototypes

Thumbnails

In order to best show the contents of a video, we employ the use of interactive thumbnails that the user can use to explore. Furthermore, In Figure 9, we show different interaction visualizations. In the top thumbnail, the interval that the thumbnail represents is shown clearly on the bottom of the thumbnail (the highlight), and placing the cursor over the thumbnail shows a seek line to indicate to the user the temporal location of the frame being previewed. This was found to be somewhat confusing to users. Instead, the bottom thumbnail is more simplified and simply shows a line indicating the thumbnail's preview location, and the highlight is removed.

To highlight important parts of the video, we move to Figure 8. The bottom thumbnail was found to be obtrusive and occluded too much of the thumbnail to be useful, even if the highlight was translucent. We instead opt for the top thumbnail instead as it provided enough information but did not occlude the thumbnail to an unusable level.

Filmstrip

We have three designs for the filmstrip, and each has a different method for representing the user's viewing behaviour; these designs can be seen in Figure 10.

In the first design, each thumbnail represents one fifth of the entire video, the highlighted parts represent the most watched video intervals. If the automatic method for thumbnail selection is chosen, the midpoint of each highlight within the thumbnails is chosen, as seen in the diagram. Like the thumbnails above, this highlight proved to occlude the contents of the thumbnail too much to be useful.

In the second design, the interval that each thumbnail represents is variable, as well as the size. These would change based on the relative view count of each interval, allowing users to quickly see which parts of the video has been seen most often, and draw the user's attention to a thumbnail they are more likely to recognize. However, when multiples of these filmstrips are shown, there is a lot of visual clutter which may make it more difficult to search for a specific video due to an overload of visual information.

Finally, the last design greatly resembles the first one, except for the visualization of the viewing behaviour. The curve shown is a histogram of the number of times each frame of the video has been viewed. This design, while highly informative, was visually unappealing and as we found out through

user feedback, provided more information than would be necessary for searching through a video.

Video Player

This is the interface of video player with function of manual thumbnail selection. There is a "star" button on the bottom right of the video player, which allows the viewer to manually select any frame as thumbnail. The selected thumbnails are displayed in the filmstrip to represent the video content, see Figure 11.

Alternative Interface Designs and Interactions

In the process of designing methods to allow users to easily find videos they have already seen, we looked at other design possibilities. In Figure 12, we show in Design Sketch 1 thumbnails being generated based on the highlights within the timeline. These thumbnails are automatically generated and are not limited to five, like in the designs discussed prior. In Design Sketch 2, we allow the user to manually create their own thumbnails through explicitly marking a video interval that they want to save and watching again. In Figure 13, Design Sketch 3 gives the user two modes for viewing the video: a recording mode, and a playing mode. Creating a thumbnail is thus accomplished by going into the record mode, and then clicking a dragging along the timeline. We found that these methods either required too much user intervention (they were unlikely to do these tasks while watching a video), or that they did not provide enough information to inform the viewer the underlying mechanism, which causes confusion and frustration.

Two Competing Methods for Video Selection

In Figure 14, we show two competing designs that we presented to users to evaluate. In this stage of the process, we had two methods for video thumbnail selection: manual and automatic. For the manual interval selection, users would click the star on the player to create a thumbnail that would appear in the video selection screen. There would be a final thumbnail to allow the user to view the entire video should they need to. The automatic thumbnail selection is shown on the right, and works as described above.

User Feedback

The design of the thumbnail, though it has its own rationale, confuses most of the first-time users. The users reported it hard to interpret the connection between the displayed thumbnails and highlighted parts.

The first design of the filmstrip, is fairly straightforward but makes the interface cluttered and has the issue of occluding the thumbnails underneath. The second design seems unorganized and visually unappealing. The third design, also suffers from occluding the thumbnails and causes confusion for first-time users.

For the video player with function of manual selection, users reported that while it was probably better suited to be able to select your own thumbnails manually, the additional work required makes it intrusive and distracting in terms of the video watching experience.

Final Design

In the final design, we outline the design decisions that made it into the implementation of the high-fidelity prototype.

Thumbnails

In the thumbnails, we chose to use the top design in Figure 8 for representation, and the bottom interaction technique in Figure 9. These provided both enough information for the user without the clutter inherent in the other designs. Thumbnails in this design also show the timestamp that the preview is showing, giving users a better sense of the temporal location of each thumbnail in the video.

Filmstrip

The design of the filmstrip in both the player and the video selection interface resembles the first sketch in Figure 10, with the exception of the highlighting. This was replaced with the strip highlighting shown in the thumbnails, which reduced the occlusion made upon the video. The final filmstrip can be seen in the bottom of Figure 1.

Video Player

The final interface of the video player incorporates both automatic and manual elements to the current interface. There is a filmstrip consisting of 5 dynamic thumbnails, each of which represents one fifth of the entire video and automatically generated by the system while watching the video. After it's done, the user can choose to alternate these thumbnails through navigating the filmstrip and clicking. Because this is done after the video watching, it requires no additional attention from the viewer thus the distraction is minimized, see Figure 1.

Video Selection Interface

Finally, we included three video selection interfaces, each allowing varying levels of detail and control. The first is the automatic method described above, which can be seen in Figure 2. This representation shows users select frames from the video that they have seen the most within the five thumbnails. It also shows the user which parts of the video they watched most from the blue bars lining the top of the video. The user can place the mouse over each thumbnail to explore the one-fifth of the video that the thumbnail represents. This gives the user a quick method to explore the entire video both at-a-glance, and in detail. The second design we implemented is the mixed method, seen in Figure 3. The design is essentially the same as the automatic method, except that it shows user-selected thumbnails, as indicated by the star in the top-left of the thumbnails. Finally, our third interface is inspired by Youtube, shows a single thumbnail that again, the user can hover the cursor over to explore the entire contents of the video, shown in Figure 4.

METHODOLOGY CONSIDERATION

In personalization, studying a decision on behalf of the participants is challenging, as extreme care must be taken to minimize any bias of that decision process. Our goal is to develop a methodology where during the study, the user can freely choose to use the manual method, namely bookmark function, or their watching behaviours can be automatically recorded to generate thumbnails that can remind them of the

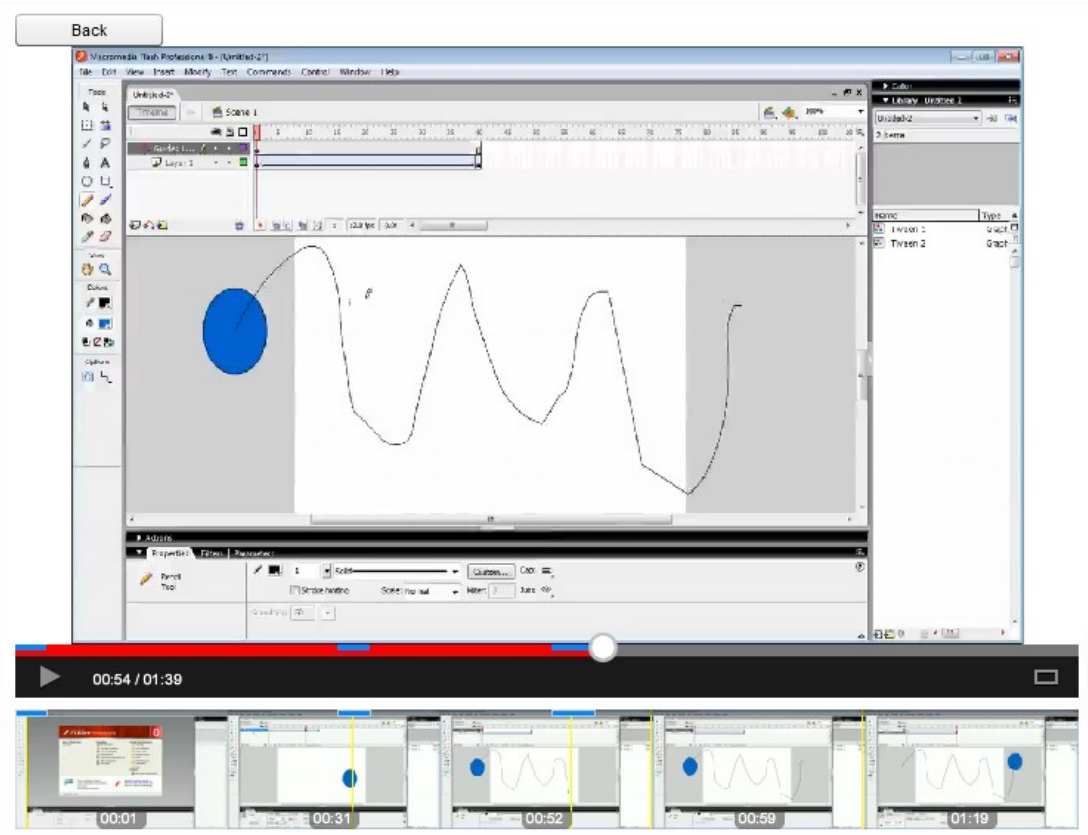

Figure 1. The video player interface: The main body is the video player itself. Below it is a seek bar that highlights the most seen intervals of the player, as well as the temporal location of the video the user is currently watching. On the bottom is a filmstrip of thumbnails displaying a preview of the video's representation on the video selection screen.

video content when searching or browsing for their previously seen videos. The primary dependent variable that we are interested in is time taken to find the watched videos. Here, we discuss the factors and process of developing our study methodology in detail.

Video Genre

We limited the scope of studying the utility of the interfaces to educational and tutorial videos. As such, we chose tutorial videos as our experiment videos for two reasons. Firstly, our research interest is if the automatic and manual methods can select thumbnails that can help users search the watched videos efficiently. So in our experiment, we chose tutorial videos and asked users to follow the tutorials, as a way to simulate real-world scenarios and intrinsically trigger the users to use playback, pause and bookmarking functions. Secondly, in order for such a study to work, there must be a reasonable motivation for the users to rewatch the videos. This means the content of the videos must be attractive enough to motivate participants to rewatch. In other words, the tutorial videos can shorten the gap between experiment and real-world scenarios for motivation to rewatch videos.

Hindrance and Interruption Simulation

Unlike real-world scenarios, where the hindrance and interruption for users to find the watched videos can propagate over long periods of time, lab experiments have a limited duration. So, to simulate the real-world hindrance and decrease the influence of participants' short memory for experiment videos on the accurate measure of time metric, we have users watch multiple videos as distractors and simulate real-world scenarios..

USER EVALUATION

To evaluate the prototypes identified above, we performed several informal evaluations along the iterative process. To start, we conducted informal interviews to gauge the utility of different methods for video selection. Through these interviews we discovered shortcomings in our designs and refined them. We then created the paper prototypes outlined above and conducted informal walkthroughs and think alouds to evaluate the interfaces. Finally, we created a high-fidelity prototype and evaluated that in a formal evaluation.

Experiment

We conducted a formal evaluation to evaluate the efficiency as well as perceptive experience of three methods for video selection: YouTube style, automatic thumbnail generation, and mixed thumbnail generation. The purpose of the experiment was to determine the speed of video selection for these three methods, the error rates, and qualitatively, which of the three methods was more desirable and easy to use. We hypothesized that:

Figure 2. The thumbnails in this filmstrip are automatically chosen by the system to show the most viewed intervals of the video. Each thumbnail represents one-fifth of the video, and by placing the cursor over the thumbnail, the preview can be changed to show the appropriate frame of the video. The blue bars on top of the thumbnail indicate the most viewed intervals of the video, and the selected frame shown is the middle frame of those intervals for the respective thumbnail.

Figure 3. The thumbnails in this filmstrip are selected in two ways: automatically, like in the Figure 2, or user selected. While in the player, the user is able to select the thumbnail that is used to represent the video in the video selection screen. This is accomplished by placing the cursor over the filmstrip in the player, hovering the cursor across the filmstrip until the desired thumbnail appears, and then clicking on the thumbnail. The selected thumbnail is then permanent and will be shown in the video selection screen. Deselecting the thumbnail and reverting back to the automatic method of thumbnail selection is accomplished by clicking the star.

H1 automatic and mixed would be significantly faster than YouTube,

H2 mixed would have significantly lower error rates than YouTube and automatic,

H3 and that users would like the automatic method more than the other two methods.

Participants

We recruited eight unpaid volunteers ranging in age from 19 to 61 (3 females). Each participant worked on the task individually. All the participants have watched tutorial videos online and revisit video that they have seen before.

Apparatus

The experimental interface was a modified version of the last prototype presented in the Prototypes section of this paper. The interface was modified to follow the experimental procedure, complete with prompts, video presentation, randomization of tasks, etc, as will be discussed in the procedure.

Procedure

(1) Introduce the participants to the interface. We first showed them the player, and introduced the filmstrip and how it keeps track of the user's watching history (the blue bar) and how the thumbnails in the filmstrip are chosen using this metric. We then showed them how to manually change the filmstrip's thumbnails, and how to reset it. Following this, we went to the video selection screen and showed them all three methods that video can be represented: automatic, mixed, YouTube. (2) We allowed the user to try the above for themselves, with another video: watching a video, skipping around to create a viewing pattern, selecting thumbnails for the filmstrip to show, and flipping between the three methods for selecting video. (3) The participants were presented three video tutorials that they were asked to follow and complete. These three videos were: making a paper snowflake [1], making a paper star [2], and creating a motion tween in Adobe Flash [3]. They were told that they would need to find these videos again in a list of videos and to use the tools provided if they saw fit. (4) The experimental interface presented the participants with a screen that instructed them to find one of the videos they previously watched using one of the three methods. After clicking a start button, the users were presented with the video selection interface, in which the users would then have to select the correct video amongst five other distractor videos. The order of these videos was randomized. This process was repeated 16 times with different combinations of videos and video selection methods. The order of the videos presented was randomized, but it was ensured that a video would not be presented twice in a row. The order of the method was fixed: automatic, mixed, YouTube. (5) After completing 16 trials, the participants were asked to complete a questionnaire which asked them to rate their general experience with the interfaces (5 point Likert scale: confusing or clear, useless or useful, difficult or easy, slow or fast, frustrating or feasible), to rank the interfaces, as well as provide comments about the three video selection methods.

RESULTS

Using a Huynh-Feldt correction for non-sphericity, a two-way repeated measures analysis of variance revealed a significant effect of method (F = 8.34, p = 0.018), no significant effect of video (F = 1.51, p = 0.295), and no significant method

[1] https://www.youtube.com/watch?v=eDctNgkNf7c
[2] https://www.youtube.com/watch?v=scdqwjgnIas
[3] https://www.youtube.com/watch?v=RD546HqP-Lo

Figure 4. The more traditional video selection method features a single thumbnail. This is enhanced from the YouTube version and allows users to hover over the thumbnail to explore the video.

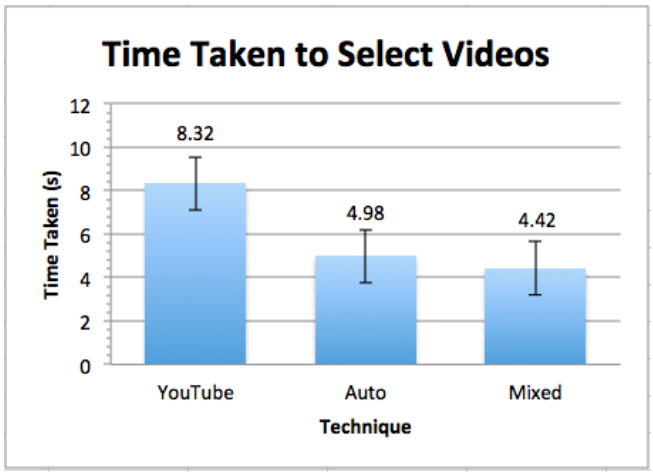

Figure 5. Average time taken to select the correct video using the specified method. Participants took significantly less time to select videos using the mixed and automatic methods over the YouTube method.

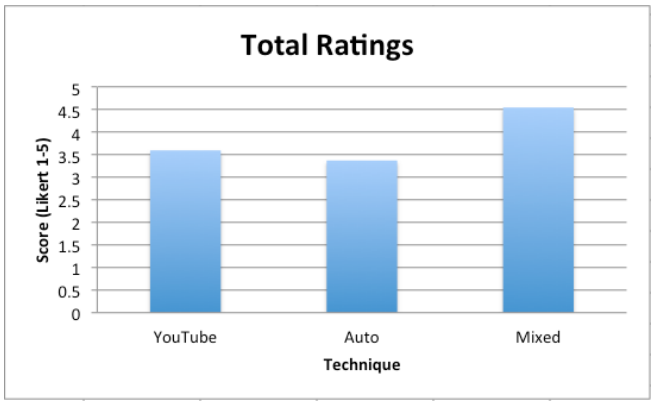

Figure 6. Participants were asked to rate the three methods on 5 metrics on a 5 point Likert scale: confusing or clear, useless or useful, difficult or easy, slow or fast, frustrating or feasible. These scores are added up and their means are graphed here.

by video interaction (F = 0.91, p = 0.537) on time taken to select videos. A follow-up pairwise comparison revealed that the automatic method performed better than the YouTube method (p = 0.015), the mixed method performed better than the Youtube method (p = 0.009), but the mixed method did not perform significantly better than the automatic method (p = 0.61). These results can be seen in Figure 5. In terms of error rates, only one participant made any errors at all, and these were distributed equally among all the techniques.

The ratings of the methods from the questionnaire revealed no significant differences between each method (F = 3.084, p = 0.067). The ratings can be seen in Figure 6. We performed a Friedman test on the rankings of the methods and found a significant difference (χ^2 = 7.00, p = 0.030). A Wilcoxon Signed Ranks Test revealed a significant rank difference between automatic and Youtube (p = 0.026), and mixed and Youtube (p = 0.033), but no difference between automatic and YouTube (p = 0.61). The final ranking is thus: (1) Mixed, (2) Automatic, (2) YouTube. These preference ranking results are shown in Figure 7.

Aligning with the quantitative results that mixed method has a significantly higher ranking than automatic (p = .033) and YouTube (p = .026), participants had positive comments on mixed method.

Included here are general comments made by our participants about the interface.

"[Mixed] gives me control over what I would like to choose to remind myself of the clip. The bookmarking idea (and the yellow symbol) really stood out and helped me recall what I just watched. Also, when I was bookmarking, I had to really think about which part of the video would offer the biggest impression of that particular video." - P1

"Automatic method can be useful when I was busy with watching and had no time to record manually. But I can use the manual method to bookmark the featured thumbnails. So the combination of manual and automatic method is quite important." - P4

Although from time taken to select videos results, YouTube method was significantly slower than the two other methods, some participants liked YouTube as first or second best. And I think what P1 said can explain this situation well.

"I like YouTube as second best because I am used to it from past experience and that User interface was neat and do not have too much going on. (My eyes can relax unlike Automatic version)." - P1

Actually, before our user evaluation we had hypothesis that the automatic method should be favored most, but the ranking results verified we were wrong. And here some comments from participants may give us some explanation.

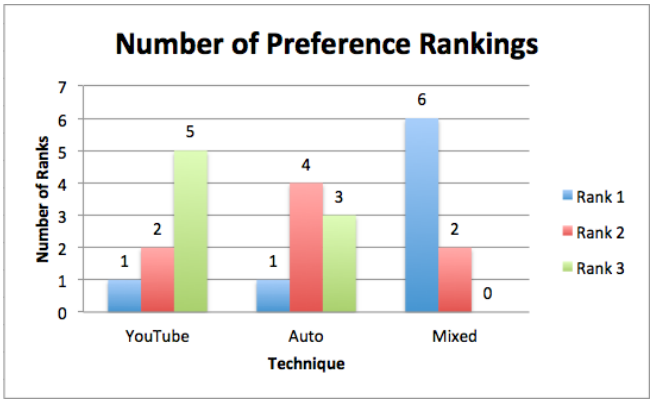

Figure 7. When asked to rank the methods, participants ranked the mixed method significantly higher than the other two methods.

"I favour Automatic the least because I couldn't really input my own preference over a piece of image that would be most useful or biggest reminder to me. What's more, there was too many images going on. Overall, I felt there was too much going on in the page (too cluttery, or overstimulating) and I got even more confused" - P1

"This looks cool, but the automatic method doesn't seem useful, and I'd like a manual method." - P7

From demographic part results in our questionnaire, P4 and P6 watched video tutorials 7+ in a month and they had some interesting comments and advice.

"there is restriction to bookmarking the thumbnails. For one interval, there may be more than one thumbnail that needs marking but another interval may have none." - P6

"Filmstrip is a great improvement vs. YouTube. Manual method can be combined with notes. And Manual method can record the featured thumbnail to help find the video in a few months. After finding the video, the blue bar [auto method] can help decide to watch from which intervals." - P4

DISCUSSION

Through our analysis, we have identified that participants spent significantly less time using both automatic and mixed method than YouTube in selecting videos. The mixed method was significantly preferred most than the other two methods. Although for participant usability evaluation, there were no significant results among three methods, these are encouraging results.

Looking back into our set of hypotheses, **H1** was found to be correct as automatic and mixed were significantly faster than YouTube. **H2** was incorrect; each method had the same error rate. **H3** was also incorrect; users in fact, liked the mixed method more than both the automatic and YouTube methods.

The analysis revealed only subtle time differences between automatic and mixed method. Larger scale studies, with more participants, may be required to measure the relative differences. As our design purpose is to help users find watched videos after a long duration in the actual application, the time-limited lab study may cannot verify our design completely. Although we have used interruptions and hindrances to simulate the long-time scenario, the short-term memory of participants for our test videos can still have a biased influence on our experiment results.

We also found that the preference ranking of participants may be related to how often they watch the tutorial videos. One participant (P5) who does not watch tutorial videos never used the bookmarking function in the whole experiment. He said he had no time to bookmark when watching videos. Coincidentally, he made the most errors in our video selection tasks among other users. But two participants who watched tutorial videos more than 7 times a month spoke highly of the manual method, and thought it would help them find the videos after a long time. To some degree, this indicates that the manual method can be quite useful in real application although it has extra operation cost.

Another important finding was that some participants preferred YouTube method over the automatic one. There are two reasons that can explain this situation. Firstly, they were already familiar with the YouTube method. Secondly, they thought the automatic method offered them less control and generated many thumbnails that were without useful information which disturbed them compared with the neat YouTube interface. But participants still commented that the filmstrip can provide more information with more thumbnails. According to the conflicts, which there are two modes to choose, namely YouTube (one thumbnail) and filmstrip (multiple thumbnails) when browsing the video list, could be potential solution. It is also interesting to note that while users had no preference ranking differences between the YouTube and automatic methods, the automatic method was still significantly faster than the YouTube method.

In summary, the combination of manual and automatic method we designed can make the video selection process more efficiently. Although they still have some shortcomings in our existing design, the underlying theory of combination of recording users' basic operation behaviors (such as playback, pause) and users' extra operation (bookmark) can be highly potential in watched video selection application.

CONCLUSIONS AND FUTURE WORK

In this paper we presented the iterative design process that we used to create engaging and useful video searching interfaces and rationalized our design strategies. We demonstrated a methodology for time-limited lab personalization study. We then used this methodology to design and execute a user study. We performed a comparative study using our proposed interfaces and the standard YouTube interface and found significant results in favour of our own interfaces. The results of the study demonstrated that our proposed automatic and manual methods can help users find videos they have seen before quickly.

In future work, we can extend our usage of mixed method to other applications, such as help users skip to video content they feel interested. And our manual method can be improved in further step, such as combination of bookmarking and taking notes to better reminder users of the video contents. Time-

limited lab studies have inherent challenges when running personalized studies that involve user choice and memory. It would be interesting to develop field study methodologies to test how closely our results would be replicated in actual usage scenarios.

REFERENCES

1. Adar, E., Teevan, J., and Dumais, S. T. Large scale analysis of web revisitation patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, ACM (New York, NY, USA, 2008), 1197–1206.
2. Al Hajri, A. *Shaping video experiences with new interface affordances*. PhD thesis, 2014.
3. Girgensohn, A., Boreczky, J., and Wilcox, L. Keyframe-based user interfaces for digital video. *Computer 34*, 9 (2001), 61–67.
4. Greenberg, S., and Witten, I. H. Adaptive personalized interfacesa question of viability. *Behaviour & Information Technology 4*, 1 (1985), 31–45.
5. Jackson, D., Nicholson, J., Stoeckigt, G., Wrobel, R., Thieme, A., and Olivier, P. Panopticon: A parallel video overview system. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, ACM (New York, NY, USA, 2013), 123–130.
6. Karasawa, K. Personalized search using identity-based hashing. In *Proceedings of the 4th ACM Workshop on Digital Identity Management*, DIM '08, ACM (New York, NY, USA, 2008), 71–80.
7. Qiu, F., and Cho, J. Automatic identification of user interest for personalized search. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, ACM (New York, NY, USA, 2006), 727–736.
8. Schoeffmann, K., Hopfgartner, F., Marques, O., Boeszoermenyi, L., and Jose, J. M. Video browsing interfaces and applications: a review. *Journal of Photonics for Energy* (2010), 018004–018004.
9. Syeda-Mahmood, T., and Ponceleon, D. Learning video browsing behavior and its application in the generation of video previews. In *Proceedings of the ninth ACM international conference on Multimedia*, ACM (2001), 119–128.

APPENDIX

Prototype Figures

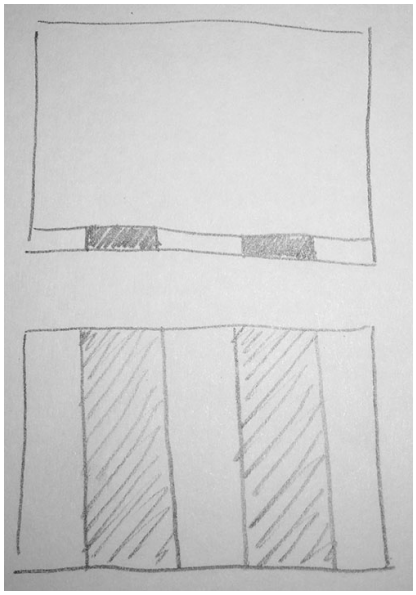

Figure 8. Two methods to represent highlights in a video thumbnail. The bottom thumbnail's highlight occluded the thumbnail too much to be useful.

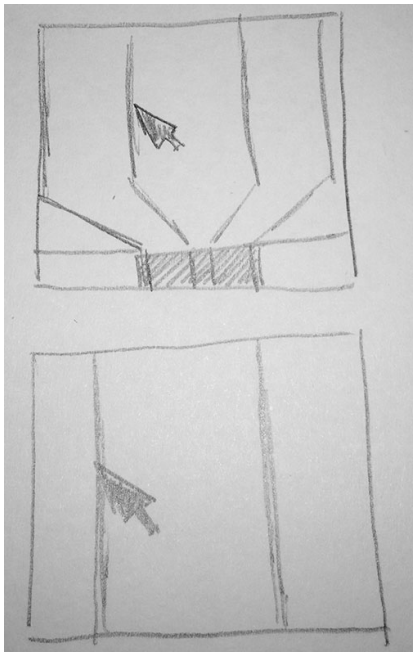

Figure 9. Two methods of visualizing interaction with the thumbnail. Here, the user can place the cursor over the thumbnail to explore the contents of the video. The top thumbnail was deemed too difficult and had too much visual clutter

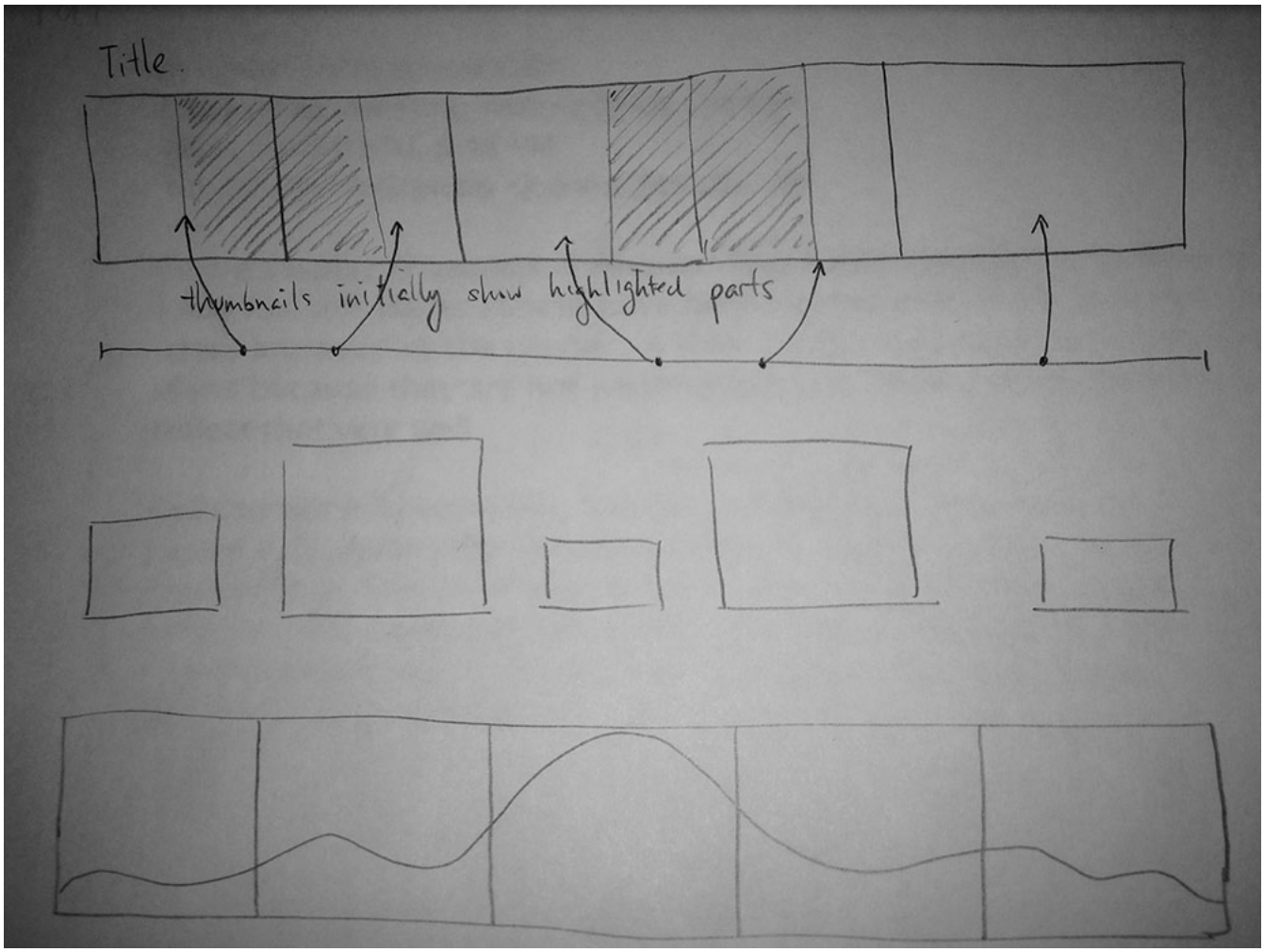

Figure 10. In the first design, each thumbnail represents one fifth of the entire video, the highlighted parts represent the most watched video intervals. If the automatic method for thumbnail selection is chosen, the midpoint of each highlight within the thumbnails is chosen, as seen in the diagram. In the second design, the interval that each thumbnail represents is variable, as well as the size. These would change based on the relative view count of each interval, allowing users to quickly see which parts of the video has been seen most often, and draw the user's attention to a thumbnail they are more likely to recognize. The last design is like the first one, except for the visualization of the viewing behaviour. The curve shown is a histogram of the number of times each frame of the video has been viewed.

Figure 11. The interface of video player with function of manual thumbnail selection. There is a "star" button on the bottom right of the video player, which allows the viewer to manually select any frame as thumbnail. The selected thumbnails are displayed in the filmstrip to represent the video content.

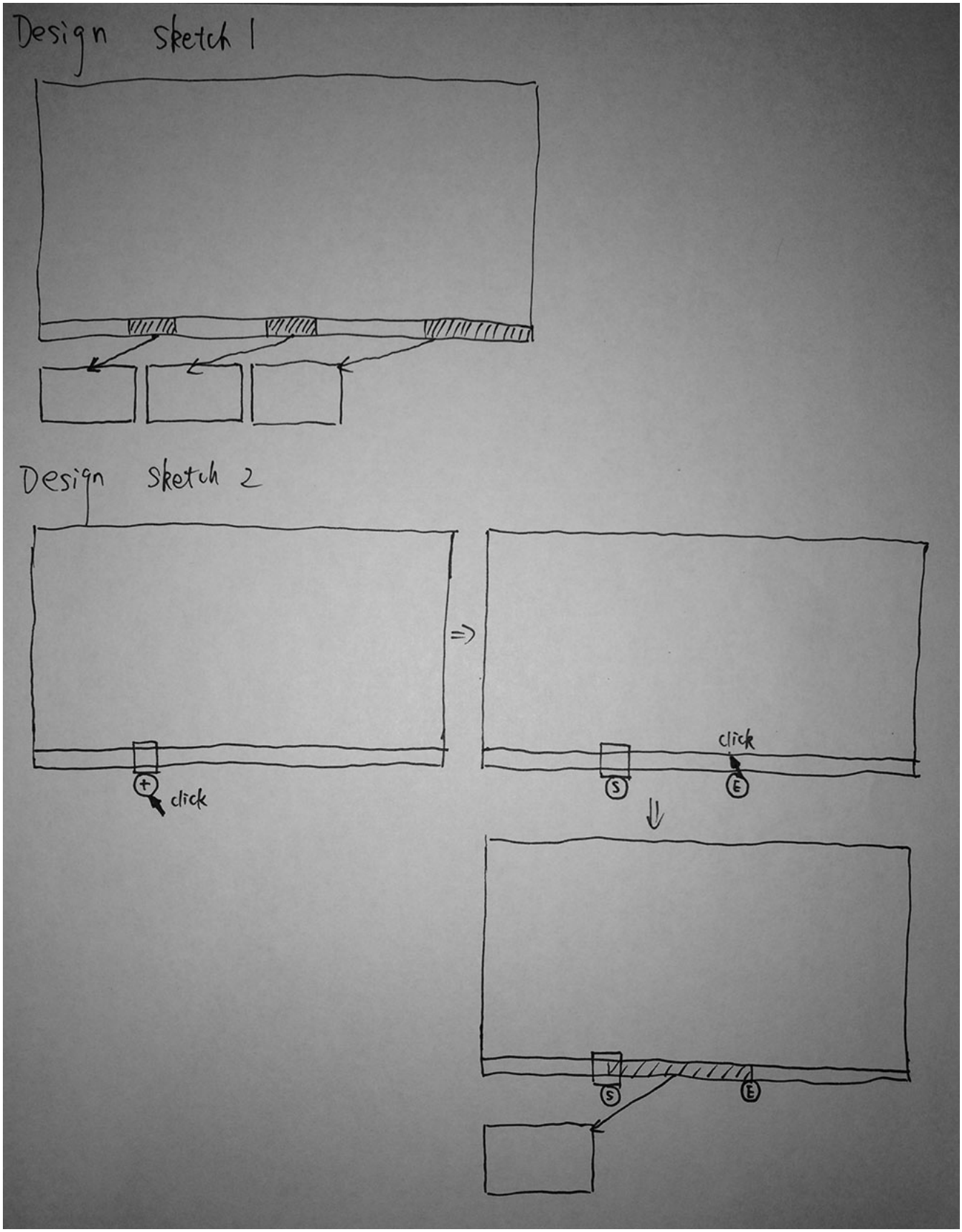

Figure 12. We show in Design Sketch 1 thumbnails being generated based on the highlights within the timeline. These thumbnails are automatically generated and are not limited to five, like in the designs discussed prior. In Design Sketch 2, we allow the user to manually create their own thumbnails through explicitly marking a video interval that they want to save and watching again. In Design Sketch 2, we allow the user to manually create their own thumbnails through explicitly marking a video interval that they want to save and watching again.

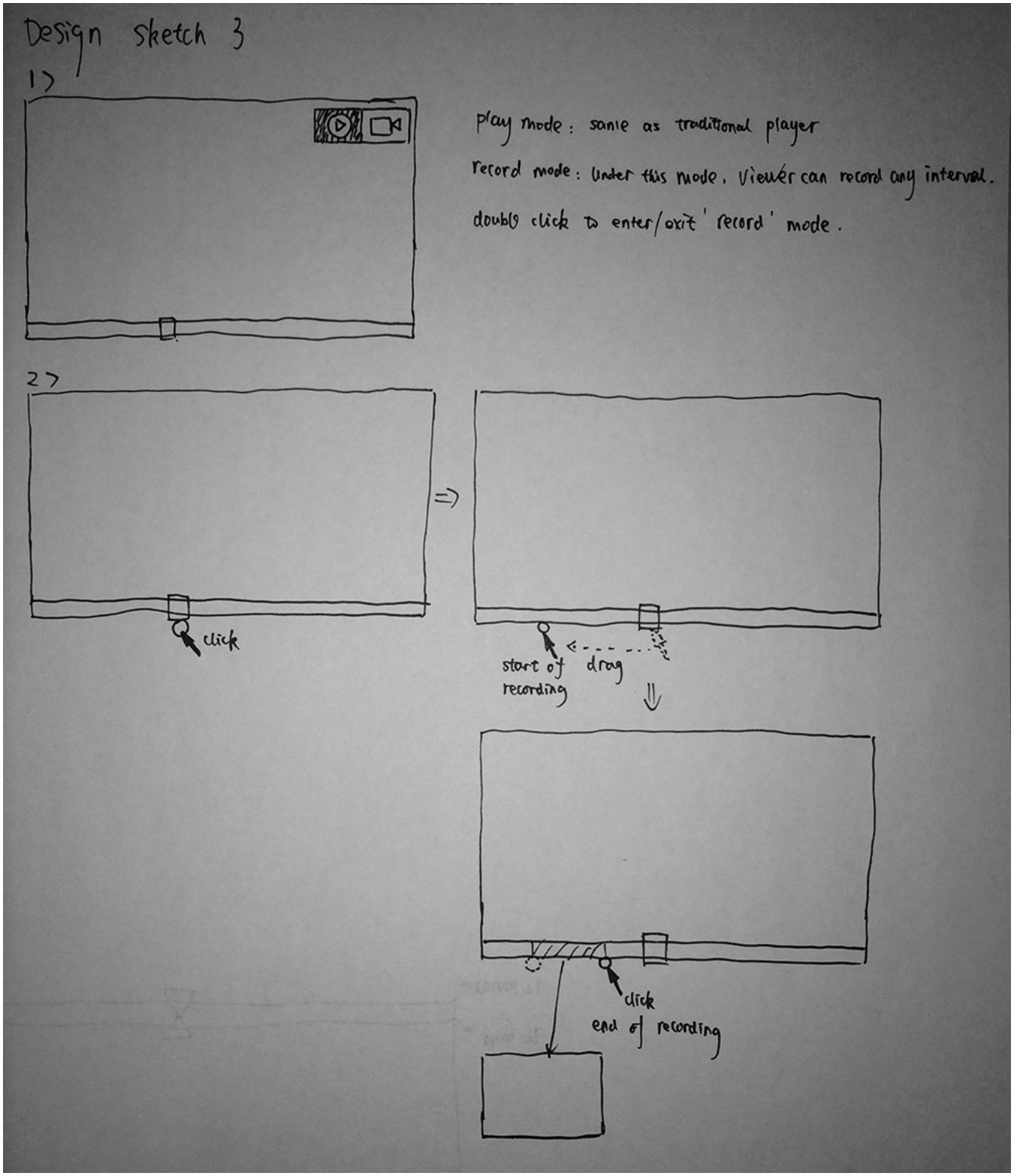

Figure 13. Design Sketch 3 gives the user two modes for viewing the video: a recording mode, and a playing mode. Creating a thumbnail is thus accomplished by going into the record mode, and then clicking a dragging along the timeline.

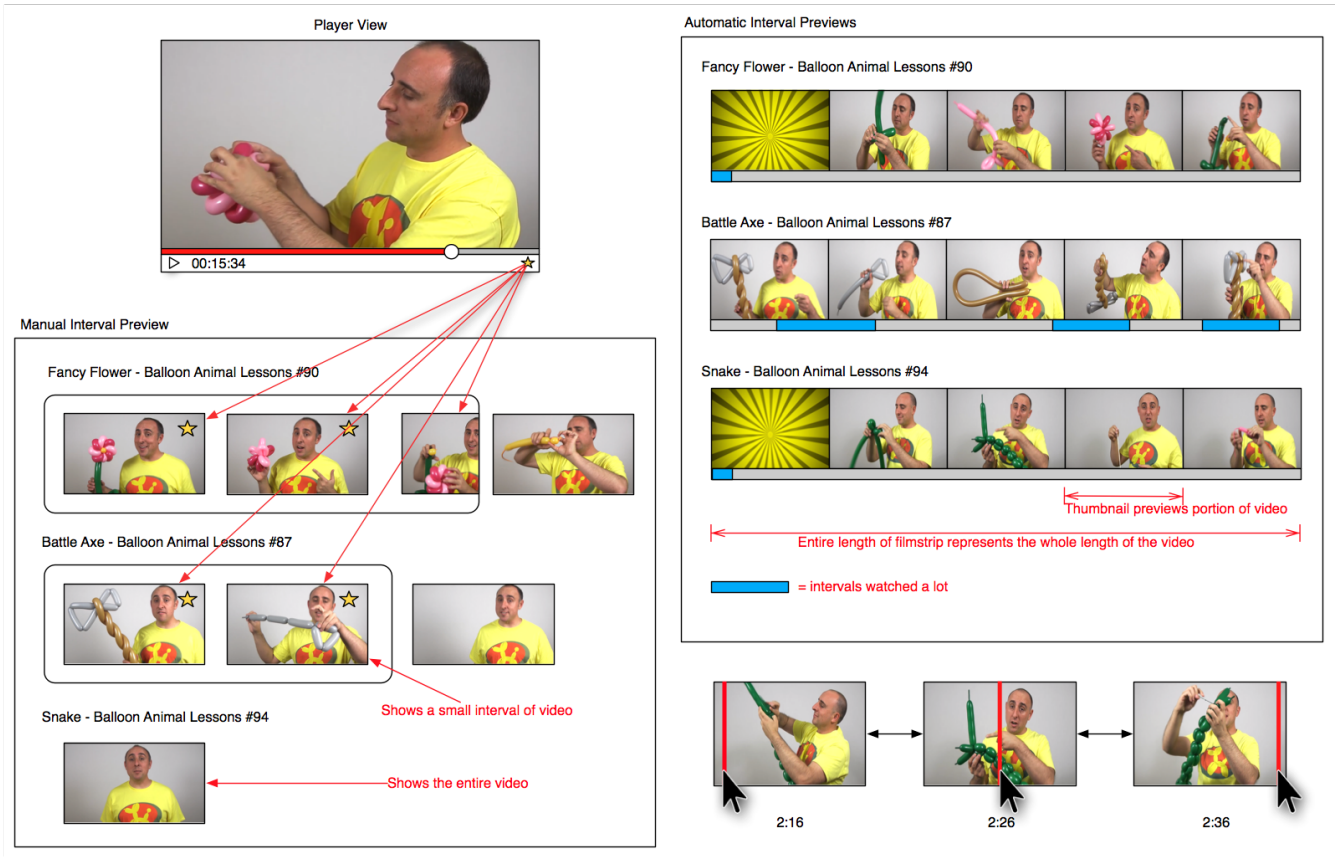

Figure 14. We show two competing designs that we presented to users to evaluate. In this stage of the process, we had two methods for video thumbnail selection: manual and automatic. For the manual interval selection, users would click the star on the player to create a thumbnail that would appear in the video selection screen. There would be a final thumbnail to allow the user to view the entire video should they need to. The automatic thumbnail selection is shown on the right.