“You’ve got video”: Increasing Clickthrough when Sharing Enterprise Video with Email

Mercan Topkara¹  Shimei Pan¹  Jennifer Lai¹  Ahmet E. Dirik²  Steve Wood¹  Jeff Boston¹  
¹IBM T. J. Watson Research Center  19 Skyline Dr, Hawthorne, NY, USA  {mtopkara, shimei, jlai, woodsp, daddyb} @us.ibm.com  
²Uludag University  Gorukle Campus, 16059, Bursa, TURKEY  edirik@uludag.edu.tr

ABSTRACT
In this Note we summarize our research on increasing the information scent of video recordings that are shared via email in a corporate setting. We compare two types of email messages for sharing recordings: the first containing basic information (e.g. title, speaker, abstract) with a link to the video; the second with the same information plus a set of video thumbnails (hyperlinked to the segments they represent), which are automatically created by video summarization technology. We report on the results of two user studies. The first one compares the quality of the set of thumbnails selected by the technology to sets selected by 31 humans. The second study examines the clickthrough rates for both email formats (with and without hyperlinked thumbnails) as well as gathering subjective feedback via survey. Results indicate that the email messages with the thumbnails drove significantly higher clickthrough rates than the messages without, even though people clicked on the main video link more frequently than the thumbnails. Survey responses show that users found the email with the thumbnail set significantly more appealing and novel.

Author Keywords
Video sharing; thumbnail view; email overload; link sharing

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors, Design

INTRODUCTION
Increasingly talks, as well as meetings, are recorded and made available after the fact as a video recording. These videos tend to be much longer than the average YouTube video, often lasting up to 60 minutes. Current methods of sharing video as an attachment or a link in an email are sub-optimal since one method has bandwidth and storage costs, and neither provide the recipient with sufficient information to determine if it is worth their time to view the recording. Corporate file-sharing systems notify users that a file has been shared with them by sending an email with a link to the file. These online systems were developed in an attempt to increase knowledge sharing, and reduce reliance on sharing files as attachments [5]. However, given the well documented phenomena of ‘email overload’ [4], users already feel that they receive and send more email than they can handle, and have developed triage methods for their email to help with the handling of the daily onslaught [3]. Thus receiving an email with a link to a video, which may or may not be of interest to the user, is most frequently ignored, unless the link is sent directly to the recipient by a trusted source [2].

In an attempt to address these issues we experimented with the use of a video processing technology we developed that allows the creator/owner of the video to share a set of automatically selected shot thumbnails via email (see Figure 1). We evaluated whether recipients of these messages would be more or less inclined to view the video, and their subjective reactions to this type of information notification. A within-subjects study showed a higher number of clickthroughs originating from emails that included the visual thumbnail set and responses to a survey indicate that users found the email with the visual thumbnail set more appealing and novel.

In the next section we briefly describe the video summarization technology, and then introduce key related work. We subsequently describe the two studies with their results. Lastly we discuss limitations of this work.

VIDEO SUMMARIZATION TECHNOLOGY
We developed a low complexity unsupervised video summarization technology for use with recordings of talks. The summary consists of salient thumbnails selected from video frames based on a ranking algorithm that takes into consideration several heuristics based on the characteristics of the domain. The thumbnails are presented in 200x150 pixel size due to the limited real estate available in emails. The technique uses the following standard steps: shot detection; shot clustering; slide detection; shot ranking; and selection of top 10 representative frames. We refer to a sequence of successive frames with similar content as a
“shot”. Shot detection is the step of segmenting video into shots. Only one representative frame from each shot is used during the clustering and ranking steps. Shot detection uses color histogram to evaluate the similarity of frames, which is standard for this technology.

Several studies show improvement in user experience when visual context is provided for web page previews (in search results or mobile web previews) [1, 6]. Aula et al. showed that users overestimated the helpfulness of a web page when they were provided only a textual snippet, and they underestimated the helpfulness when given only visual thumbnail of the page. The best judgment was achieved when both were available. Our findings also suggest that users found it more useful when visual thumbnails were added to the general information about the video.

Song et al.’s work focuses directly on the usefulness of audio and visual ‘surrogates’ for deciding whether to download a video [8]. They used a web interface to provide the users with three different types of surrogates: audio (spoken) description, a visual storyboard and both. Users performed best on the sense-making tasks when both surrogates were available. By contrast, in our work we focused on the effect of an automatically created visual summary together with manually created basic information.

### USER STUDY 1: VALIDITY OF SUMMARIZATION TECHNOLOGY

The goal of this study was to verify that the technology used to select the 10 thumbnails sent in the email did a sufficient job to represent the video. To measure this we compared the set of thumbnails selected by the technology with sets selected by 31 humans for the same videos.

#### Setup

We provided representative shots from four videos of recorded talks (Video 1: 91 shots; Video 2: 147 shots; Video 3: 182 shots; Video 4: 192 shots;) in a web interface and asked participants to select 10 shots that represented the video and would give another person an idea of what the talk was about. The frames were presented as 200 x 150 pixel thumbnails. Participants ranged in age from mid 20s to mid 50s from different ethnic backgrounds, all working in software technology at a large research company.

#### Results

We calculated the agreement of each participant with the other study participants as well as the agreement of the system’s selection with that of all participants. The agreement value is calculated by performing 1 by N color histogram similarity comparison on each of the thumbnails in the set. As shown in Table 1 there was a high variance in user agreement. For example, for video 1, when comparing sets selected by participants, there was only 65.9% agreement among participants. This result supports other research that reported that human participants do not agree strongly on a common summary [7, 9].

We modeled our evaluation on TRECVID (a video summarization benchmarking effort), which has recently started using a similar process of creating a ground truth

<table>
<thead>
<tr>
<th>System’s Agreement</th>
<th>Video1</th>
<th>Video2</th>
<th>Video3</th>
<th>Video4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average User Agreement</td>
<td>65.9%</td>
<td>76.7%</td>
<td>43.5%</td>
<td>59.2%</td>
</tr>
<tr>
<td>STD of User Agreement</td>
<td>0.58</td>
<td>0.62</td>
<td>1.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 1. System Agreement with Users

---

Figure 1. An email with thumbnails

We used several heuristics in ranking the representative frames. A cluster’s rank is directly correlated with the total shot duration in that cluster. Also, a cluster with a higher information value (i.e., different from the rest of the clusters) is ranked higher. The first shot in a slide cluster shots is selected as representative for that cluster, which aims to catch the title slide or the first slide of a new discussion point. We perform OCR analytics on all slide shots, and the number of characters extracted from a slide shot is inversely correlated with its rank (to provide thumbnails with readable text on them). Slides with greater graphic and visual appeal measures, as computed by intensity histograms, are ranked higher.

Once representative frames from the top ranked non-slide-clusters are added to the summary, slide frames from the top ranked slide-clusters are selected based on their ranking within the cluster. After an initial set of frames is selected, the diversity among these frames is re-evaluated. Low ranking frames are swapped with the next top ranked frame in the cluster until the similarity among the selected frames in the set is below a threshold value.

---

**RELATED WORK**

Given the nature of Note, we will contain our references to just the work focused on the effect of previews on content sharing.

Several studies show improvement in user experience when visual context is provided for web page previews (in search results or mobile web previews) [1, 6]. Aula et al. showed that users overestimated the helpfulness of a web page when they were provided only a textual snippet, and they underestimated the helpfulness when given only visual thumbnail of the page. The best judgment was achieved when both were available. Our findings also suggest that
with 5 trained annotators analyzing 8 hrs of BBC news videos. This truth is then used in evaluating the video summarization techniques [11].

The variance in participants’ agreement increased as the number of total shots increased. Yet in 3 out of 4 videos the technology’s agreement with the participants’ selections is within the standard deviation, indicating that the system did as well as any single participant. Only in Video 4, the system’s selection was outside of the standard deviation, which was due to the nature of this particular talk that included a review of several projects at a fast pace with only screen shots. As a result, for Study 2, we only used videos of talks that included speaker, screen shots (i.e. slides or demo) and audience shots. This was the same nature as videos 1 through 3.

USER STUDY 2: SHARING VIDEO VIA EMAIL

The focus of this research is on increasing the informativeness of the email when sharing a link to a video. Part of the goal is to support knowledge sharing within the enterprise, since if I share a video with you and you never follow the link because you can’t invest the time to see if it is truly relevant, there is a potential missed opportunity cost.

Our hypothesis was that when people can view the summary thumbnails they will be more inclined to take the time to view the video, or segments of the video, since the information scent would be higher. This hypothesis is supported by the feature integration theory in human visual attention [10], which indicates that color and intensity are primitive features and are included in the fast and pre-attentive feature search.

Our assumption was that all videos selected for the study have the same relative degree of interest to the target population since they were talks of general interest to the HCI community and the distribution list was made up of 428 people who had registered an interest in this topic.

We performed a within-subject study that involved sending the distribution list four email messages. Recipients were randomly assigned to either Group 1 or Group 2. The four videos were all similar including views of the speaker, his/her slides and the audience. The durations ranged from 59 minutes to 71 minutes.

We prepared two different types of emails for each of the four videos. The first type (WoMT – with only main thumbnail) included a main thumbnail from the video (first non-blank keyframe automatically selected by the video processing system), a link to the video, the title of the talk, the name of the speaker, 2-3 sentence abstract of the talk that was manually created by the content owner, and its duration. The second email type (WTS – with thumbnail summary) included exactly the same information as the first type, plus the automatic thumbnail summary. The thumbnails were hyperlinked to the moment in time they represented in the video. In both types of messages, clicking on the link to the video started playing the video in a new window at the start of the video recording. For the messages in WTS format, clicking on any of the thumbnails started playing the video at that moment in time of the recording.

We added the main thumbnail to both types of emails to minimize any bias based solely on the fact that one format was text only and the other was text plus image. In this way, each format had some visual element, but the WTS format had more. We sent out each email at the same time to the two groups. The emails were sent on Tuesdays and Thursdays for 2 weeks. We alternated the type of email each group received in consecutive emails. Group 1 received information about Video A in WoMT format, Video B in WTS format, Video C in WoMT format, and Video D in WTS format. This sequence of email types was reversed with Group 2.

On the third week, we sent out a survey to both groups to solicit subjective opinions in addition to the clickthrough data resulting from the email messages.

Results

In total there were 175 clickthroughs to video playback originating from these emails. 112 of these clicks came from WTS type emails and 63 came from WoMT type emails. We performed a paired t test and the mean difference between the WTS and WoMT conditions (mean WoMT = 28, WTS = 15.75) is statistically significant ($p=0.037$).

Additionally we got 27 responses to the survey. The survey questions include eleven agreement questions on the usability and attractiveness for each of the two types of emails (boring, complicated, confusing, novel, easy to use, appealing, error-prone, frustrating, intuitive, quick and satisfying). We also asked seven questions on how users interacted with the emails they received (e.g. whether a thumbnail was clicked and what would constitute a useful thumbnail). Lastly we included five general agreement questions on the usability of the thumbnails (e.g. thumbnails are useful, informative, easy to use).

Our click-through data showed that most of the recipients preferred to click on the hyperlinked title to play back the video from the beginning in both types of emails. Several survey respondents mentioned that they needed more context (in addition to the visual thumbnails) about the segments in order to decide if they wanted to start watching from that segment. This feedback is also supported by the work presented in [1], where users judged the ‘helpfulness’ of a webpage better when the preview included both a textual and a visual snippet as compared to having only one of these types.

Respondents’ top choices for the type of thumbnails they would like to see in a summary were: representative of long
segments in the video (N=13), ‘with readable text’ feature (N=10) and with ‘graphics and charts’ (N=7).

We used a 9-point Likert scale for all the agreement questions. In general, a majority of respondents favored the usefulness and ease-of-use of the emails with the thumbnail set. The mean response to judging the usefulness of email with thumbnail set was 6.08, when compared to neutral with one sample t-test we get t=2.2, p=0.038. When asked whether users preferred to receive just a link and a two-sentence summary, a high number of the respondents (16) gave a score of less than 5 (indicating disagreement).

We also provided two tables with an identical list of adjectives to measure respondents’ agreement with the descriptors for each of the email formats. Respondents found the email with thumbnail summary (WTS) significantly less boring, more novel and more appealing than the WoMT format.

<table>
<thead>
<tr>
<th>Measure</th>
<th>T</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boring</td>
<td>-3.242</td>
<td>17</td>
<td>0.005*</td>
</tr>
<tr>
<td>Complicated</td>
<td>1.531</td>
<td>18</td>
<td>0.143</td>
</tr>
<tr>
<td>Confusing</td>
<td>-0.230</td>
<td>18</td>
<td>0.820</td>
</tr>
<tr>
<td>Novel</td>
<td>6.364</td>
<td>18</td>
<td>0.000*</td>
</tr>
<tr>
<td>Easy-to-use</td>
<td>0.473</td>
<td>18</td>
<td>0.642</td>
</tr>
<tr>
<td>Appealing</td>
<td>2.122</td>
<td>18</td>
<td>0.048*</td>
</tr>
<tr>
<td>Error-prone</td>
<td>-0.470</td>
<td>18</td>
<td>0.644</td>
</tr>
<tr>
<td>Frustrating</td>
<td>-0.087</td>
<td>18</td>
<td>0.932</td>
</tr>
<tr>
<td>Intuitive</td>
<td>2.044</td>
<td>18</td>
<td>0.056</td>
</tr>
<tr>
<td>Quick</td>
<td>0.829</td>
<td>18</td>
<td>0.418</td>
</tr>
<tr>
<td>Satisfying</td>
<td>1.019</td>
<td>18</td>
<td>0.322</td>
</tr>
</tbody>
</table>

Table 2. Survey responses to two e-mail formats (WTS vs WoMT) using paired t-tests.

Several users mentioned that while they did not clickthrough to view the video, they saved these emails in a folder to view at a later time. They mentioned that these thumbnails would also provide a mechanism to recall which talk had the information they were looking for. One of them mentioned that the images helped sell the content of the talk whereas an attached powerpoint requires effort to load and view in order to decide if the talk is interesting.

LIMITATIONS OF THE STUDY
We feel this is preliminary work that indicates a promising direction for increasing the information value of messages when sharing video content. A limitation of the finding is that like many studies, there is a self selecting population that responds to either one of the messages, or the survey. The very busy people (or people with different email triage methods) did neither. Also, a finding that the interface with the thumbnails is significantly more novel should not be surprising, since this email type was compared with a more traditional type containing a link and descriptive text. However, of note is the finding that the email format with thumbnails was significantly more appealing, and that it drove significantly more clickthroughs. With the increased use of video in organizations it is key to understand how to best share the content contained in them.

We would like to gather additional data on the thumbnails (e.g. which types, if any, result in higher clickthrough) as well as test different layouts for the email message. Some respondents indicated that they clicked on the main link rather than the thumbnails since they did not realize the latter were clickable. Even though the message stated this, and each thumbnail had a blue colored link under it, this must have been missed as users quickly scanned the content of the message.

ACKNOWLEDGMENTS
We thank the users for their participation in our studies and the anonymous reviewers for insightful comments.

REFERENCES
4. N. Ducheneaut and V. Bellotti. E-mail as habitat: an exploration of embedded personal information management. Interactions 8, 5 (September 2001), 30-38.