Calculation of an Aggregated Level of Interest Function for Recorded Events

Rahul Nair
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
rahulnair@acm.org

ABSTRACT

As recording technology becomes pervasive there is a dramatic increase in the number of events being recorded in multimedia. The challenge now facing users is to quickly view the recorded content in the least amount of time. While there are several methods to analyze video based on ambient noise, scene changes, slide transitions, etc..., these techniques merely find features in the recording, they do not reveal which sections are important.

This paper presents a method to calculate a Level Of Interest (LOI) function for an event by aggregating bookmarks made by the event attendees. The findings of a preliminary evaluation of the LOI function are also presented along with the design of an LOI based video browser.

Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing – *Indexing methods*; H.5.2 [Information Interfaces and Presentation]: User Interfaces – *User-centered design, prototyping*;

General Terms

Algorithms, Design, Measurement, Experimentation

Keywords

Level of Interest, visualization, bookmark aggregation, multimedia, video browsing, skimming.

1. INTRODUCTION

As recording technology becomes cheaper, better and easier to use we are facing a deluge of multimedia recordings. Several companies have begun to record the content of internal seminars as well as weekly staff meetings. Universities are recording the content of their lectures for both distance learning and archival purposes. As the amount of recorded video begins to grow it is becoming increasingly difficult for users to find the time to actually watch the recordings.

MM'04, October 10–16, 2004, New York, New York, USA.

Time compression of audio-video [1] has allowed a 1.5 to 2.5 fold speed up but needs greater concentration from the user while still making him/her skim the entire video.

Another method is to use various automated feature detection algorithms to demarcate episodes in the video [3] as well as to automatically summarize it [4]. Techniques like applause detection, slide transitions, shot boundaries, speaker changes, etc..., allow users to quickly find various sections of the video while still requiring that the user watch each section to decide if it is important. In the case of auto summarization methods the user is no longer in control of which sections of the video to watch.

Manually created transcripts and summaries are usually better than automated methods [4] but are too expensive to be used on everyday recordings. Since video transcription is often outsourced there is an additional question of information security.

This paper presents a new technique which uses input from the event attendees to determine the important sections of a video. The results of a preliminary evaluation of the validity of the LOI function are also presented.

2. INPUT METHODS

Since most automated analysis techniques for video still require some amount of human input, this research was begun with the premise that only a human being would be able to correctly rate the importance of a particular section of the meeting. The first attempt at calculating the importance of various sections of a meeting was based on the assumption that attendees take notes only when a particular section interests them.

The initial prototype (Figure 1a) required the meeting attendees to use a handheld barcode scanner to scan a barcode when they took notes. These barcodes were printed on special sheets of note paper that were given to the user and contained several different barcodes corresponding to the possible types of notes (e.g.: point, question, note, etc...). Once the user wrote a note, they scanned the barcode corresponding to the type of note. Users were also encouraged to scan barcodes and create bookmarks of important events even if they did not want to take notes. The barcode scanner automatically time stamped the scanned barcodes and stored them in its onboard memory. At the end of the presentation the users notes were manually transcribed and associated with the appropriate time stamps. The time stamps were then used to construct the LOI curve which is described in later sections.

User feedback on the initial prototype suggested some problems with the design. Users often forgot to scan barcodes after writing their notes. Some users also wished to have the additional functionality of being able to take notes in the past (i.e.: "This note is for the topic discussed 5 minutes ago"). The prototype also required a high degree of input from the experimenters to transcribe notes, collect barcode readers, synchronize timestamps, etc... Based on the above considerations, a prototype note-taking client application (Figure 1c) was designed for meeting attendees with laptop computers. With the new client users could directly take notes on their laptops; once the note was complete they pressed a button to send the note to a central server which time stamped and identified them. The options available to the user included setting the timestamp of a bookmark to the past, marking the note as private so that the text of the note would be available only to them and selecting the type of note. As with the previous prototype the users were encouraged to make bookmarks without any notes if they felt that a particular section was especially interesting.

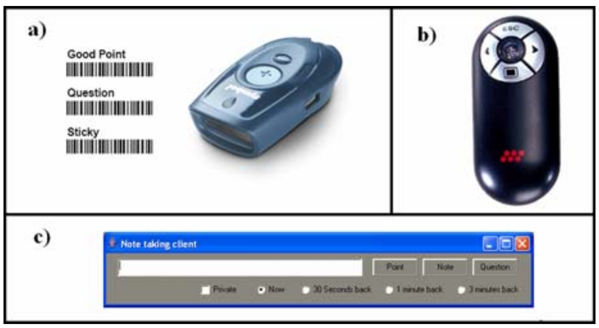

Figure 1. Input prototypes: a) Handheld barcode scanner; b) Wireless "clicker" device; c) Laptop client application

While the second prototype was more successful, users did not want to bring a laptop to a presentation solely to take notes. Users also complained about the high cognitive load of typing notes and selecting the type of note during the meeting. This feedback led to a third prototype of a wireless "clicker" device (Figure 1b). The "clicker" had a set of buttons on it, which the user could press in order to bookmark a specific point in the presentation. When a user clicked the device it sent a signal to the clicker control computer. The computer then sent a bookmark to the bookmark/note collecting server which time stamped and stored it. Users were very comfortable using this prototype since they could use the system along with their traditional paper and pencil notes with no difficulty. They were also pleased that a laptop was not required and that they no longer had to select a type of note/bookmark.

As an incentive for users to use the different prototypes, all users were provided with a private online version of the presentation recording. This private version consisted of a video of the presentation and was synchronized with both the users' own bookmarks/notes as well as the presenters slides; this allowed the navigation of the recording based on bookmarks. The users were also automatically presented with a public group recording which contained an LOI function that was created from the public notes/bookmarks of all the users put together. This allowed the users to better understand the purpose of the system as well as letting them directly experience the end results.

3. LOI CALCULATION

The bookmarks/notes obtained from the users are a one dimensional data set (Figure 2a) which gives an instantaneous interest peak at that point. However in order for the raw data to be useful it must be translated it into an interest score for that instant. This score is further converted into the LOI curve which gives the user a graphical representation of the meeting attendees' interest in the talk. This LOI curve is a simple two dimensional graph with time on one axis and the calculated interest values on the other.

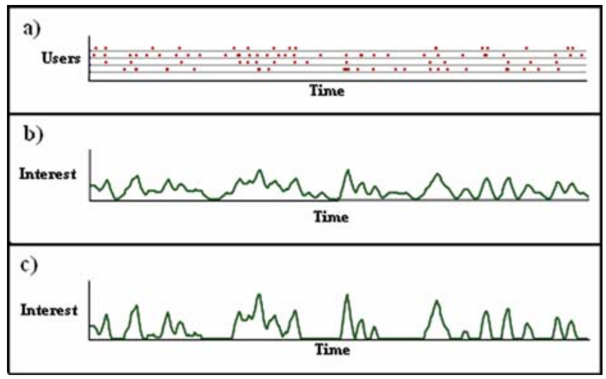

Figure 2. Calculation of LOI curve: a) Raw bookmark data from 4 users; b) Intermediate LOI curve generated using the sliding window algorithm; c) Final LOI curve after smoothening and normalization.

In order to create a continuous curve, a sliding window algorithm is used to calculate the effective level of interest (Figure 2b) at an instant. Empirical tests showed that a 2.5 minute window provided a good balance between aggregating the data and forming visually distinct peaks. Usage tests also showed that users would either mark an event as important at the very beginning (predicting that the next few minutes would be important) or more commonly, bookmark the event after it ended. This led us to use an asymmetrical window which used bookmarks 90 seconds after the point and 60 seconds prior to it. Using this technique pushed the peaks of the LOI curve earlier in time and simplified navigation since the peaks became closer to the beginning of the events.

The height of the LOI curve at a point depends not only on the number of bookmarks present in the sliding window, but also the timestamp of each bookmark in relation to the point. The weighting of each bookmark is inversely proportional to the time difference between bookmark and the time represented by the point on the curve. This is done to create a smooth curve as well as to create sharper peaks when there is a cluster of bookmarks.

A final transformation (Figure 2c) is applied to the LOI curve in order to smooth out the fluctuations in the curve. The standard deviation of the entire LOI curve is calculated and used as a threshold value; the curve is then rescaled such that only the points that exceed the threshold value will be visible in the LOI curve. This reduces the number of fluctuations in the LOI curve thereby accentuating the main peaks. In this stage the curve is also normalized so that the maximum height of an LOI peak will remain within set limits. This also helps create uniform curves irrespective of the number of bookmarks and users.

4. EVALUATION

The prototypes were used to record data during 6 presentations ranging from less than 50 minutes in length to almost 2.5 hours. The number of users per session varied from 2 to 8 and used any one of the 3 prototypes described above.

The collected data contained 359 bookmarks from 23 people (10 unique users) over a total duration of 7 hours. The average time between bookmarks varied from 28.7 seconds to 175.6 seconds, while the number of bookmarks per person per presentation ranged from 10 to 25.5. There were 64 instances where users made bookmarks within 10 seconds of each other and 26 instances of users marking the same spot twice. The latter behavior is probably due to users making multiple bookmarks to emphasize the point.

Another interesting feature was a type of social clustering of bookmarks where users with different backgrounds marked different sections as interesting. While the bulk of the bookmarks were made in large clusters around specific points, there were occasional outliers and mini-clusters that were often created by users who had a different background (and usually different interests) than the majority. It was also noticeable that certain users' bookmarks were correlated to a high degree. This clustering effect could be exploited to customize LOI curves for each user by only using bookmarks made by people with similar interests.

The LOI curves calculated from collected data had peaks over an average of 43.9% of their length with a standard deviation of 13%. After calculating the LOI curves, the video was manually browsed to extract the topic being discussed in each of the LOI peaks.

In order to test the validity of the LOI curve, a questionnaire was prepared for the presentation attendees. The questionnaire first asked them to list what they felt were the most important points brought up in the presentation. After they filled this they were presented with the list of topics manually extracted from the LOI curve and asked to rate those points on a scale of 1-5, with 5 being very important and 1 being unimportant. The final section of the questionnaire asked users if the list had either missed any points or if it had reminded them of points that they had forgotten.

A total of 23 responses covering 5 talks were collected, the questionnaire respondents were all members of various research groups who attended the presentations and in some cases had provided the input notes/bookmarks for the LOI calculation. Since the questionnaires were conducted a few weeks after the actual presentations some of the respondents were unable to answer all the questions.

The average score of the points extracted from the LOI curve was 3.43 with a standard deviation of 1. Of the 93 points listed, 27 received an average rating of 4 and above, indicating that a large number of the points are quite important. This indicates that the LOI curve is very good at identifying relatively important points.

Another metric used was asking the users if the list reminded them of points that they had forgotten. Subjects unanimously agreed that every list reminded them of points that they had forgotten. A few subjects said they had forgotten most of the presentation and that the list quickly reminded them of outline of the presentation. The respondents listed a total of 77 specific points that they were reminded of. Significantly the average score given by users to the points that they had forgotten was 3.76

When asked if the lists had missed any point only 4 users felt that points had been missed. They listed 5 different points as missing, of which two points were represented as peaks in the LOI but were not explicitly mentioned in the list since they were supporting arguments to a larger point. In no instance did two users report the same point as missing which indicates that some of the missed points could be an example of the social clustering effect. Since the LOI curve follows the interests of the majority of the bookmarking users it is quite possible to miss points that are important only to a small minority. The relative lack of missed points indicates that the LOI curve is an accurate barometer of user interest.

The subjects were also asked to subjectively rate the length of the list. 19 respondents said that the number of points was "just right" with only 4 responses disagreeing. Of the 4, 2 said that a specific list was too long while the other 2 said that a list was too short.

The subjects had previously said that that they had hardly ever watched recorded presentations despite the company having an archive of over 200 recorded meetings and presentations. Only one user said that he had ever watched a recording of a presentation that he had missed. When asked if they would be more inclined to review presentation videos if they had the important sections marked (Likert scale from 1 to 5 with 1 = very unlikely and 5 = very likely) the average score was a 4.

5. RELATED WORK

There has been a significant amount of work that attempts to visualize a multimedia stream in order to enhance browsing. Foote et al [2] and Girgensohn et al [3] built a complete media-browsing platform that allowed users to view graphical representations of several different automatic analysis techniques. The user could vary the confidence level of the analysis techniques and thereby control the granularity of the visualization. However this system could merely track changes in a meeting and detect features like applause and speaker changes, it had no way to tell if the section of the video was indeed important. The determination of the importance of a section was left to the user who still had to browse a large section of the video in order to make certain that he/she did not miss any important events.

He et al [4] used a combination of slide transitions, audio pitch analysis and end user data to automatically summarize presentations. The end user data was collected by logging the access patterns of several hundred users to see which sections of the video they watched. This data was used to assign importance scores to slides based on the relative number of users per slide as well as how long users watched the video associated with a particular slide. However this system will find it difficult to differentiate between short segments of interest and uninteresting segments skipped by end users. Several users also complained that the summaries created by this method were choppy and incoherent.

Li et al [5] used standard techniques of pause removal and time compression to increase the amount of video that can be seen in a given time. They also used shot boundary detection to allow users to quickly find specific shots. While users were able to attain an average playback speed of over 140%, they still had to watch/skim the entire video to find sections of interest. This will be especially difficult if there is a single shot of a user talking for long periods.

Finally, certain polling and market research firms use hardware based methods (sliders, knobs, dials, etc...) to collect data from an audience, however these methods have not been published or described in detail. Furthermore they usually require a paid/volunteer audience to provide constant feedback while the LOI system uses a normal audience providing implicit input (via notes) or marking points of interest.

6. APPLICATIONS

6.1 LOI Browser

The evaluation of the LOI curve showed that using bookmark data from meeting attendees is very simple way of getting reliable information about a meeting. A video browser application (Figure 3) similar to [2], [3] was built to see if the LOI curve is useful aid in navigating through recorded presentations.

Users can go to any point in the presentation by directly clicking on the LOI curve which is used as a basic timeline, marking out user bookmarks and slide transitions. The LOI browser also allows users to manipulate the LOI curve by selecting the users whose data will be used to create the LOI curve. This allows users to select a specific subset of users whose interests may match their own. Additionally the users can select the types of bookmarks that will be used to create the LOI curve. One example of its use would be to create an LOI curve to find all the interesting questions that were asked during a presentation. Other methods of navigating the video include selecting a slide on the slide thumbnail strip and double clicking on specific user notes.

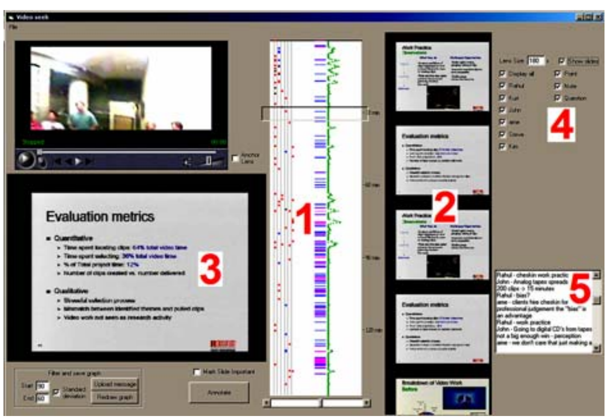

Figure 3. LOI Browser: 1) LOI curve, bookmarks and slide transitions; 2) Thumbnail strip of slides; 3) Slide synchronized to video; 4) Curve filtering options; 5)Full text of user notes

While the formal evaluation of the LOI browser interface is still pending, it has been informally tested by several users. Feedback indicates that the LOI curve is a useful method to browse a video. It was found to be especially useful while browsing through long segments without a slide change. When users were asked to skim a meeting at high speed (30 minutes for a 60 minute video) they navigated almost exclusively using the LOI curve with the slides primarily being used to recall the sections that have already been reviewed.

6.2 Other Applications

Apart from being used as a method to navigate video, the data from the LOI curve can be used in several different ways such as

- Auto summarization of the video based on the LOI peaks
- Using the LOI curve to improve the accuracy of keyframe and slide selection
- Filtering the input bookmarks to suit particular interests

While the system was designed to collect and use live data from meeting attendees it is quite easy to obtain the same information in a post hoc manner from users who watch the video recordings. Since the bookmarks (both live and post hoc) are made individuals who actually watch the presentation they have a very high probability of being accurate indicators of importance.

The required bookmark data can also be collected implicitly by adding plug-ins to note taking software to automatically create a bookmark every time the user makes a note. This will allow the system to collect data without requiring any active participation on the users' behalf.

7. CONCLUSION

This paper presents a technique of using bookmarks from attendees to create an accurate indicator of user interest level during an event. Results of the evaluation of the LOI function show that the peaks cover a small section of the video (<44%) but include all the important points. Finally an LOI based video browser was presented as a sample application for the LOI curve.

8. ACKNOWLEDGEMENTS

This research was conducted while the author was an intern with the Shared Media Group at Ricoh Innovations, Inc.

The author would like to thank the members Shared Media Group and the Multimedia Document Analysis Group for their ideas, support and feedback during this research.

9. REFERENCES

1. Arons. B. (1999) SpeechSkimmer: A System for Interactively Skimming Recorded Speech. *ACM Transactions on Computer-Human Interaction* ,4 ,1 ,1999. pp.3—38
2. Foote. J., Boreczky, J., Girgensohn, A. and Wilcox, L. (1999) An Intelligent Media Browser using Automatic Multimodal Analysis. *Proc. MULTIMEDIA '98, ACM Press, 1998, pp. 375-380.*, September 14, 1998
3. Girgensohn, A., Boreczky, J., Wilcox, L and Foote, J. (1999). Facilitating Video Access by Visualizing Automatic Analysis. *Proc. Human-Computer Interaction (INTERACT) 99*, IOS Press, pp. 205-212.
4. He, L., Sanocki, E., Gupta, A., & Grudin, J., (1999). Auto-summarization of audio-video presentations. In *Proc. ACM Multimedia'99.* ACM. pp. 489-498
5. Li, F.C., Gupta, A., Sanocki, E., He, L. & Rui Y. (2000). Browsing digital video. *Proc. CHI'00*, ACM, Pages 169 – 176.