

Assessment of End-User Response to Sports Highlights Extraction for Personal Video Recorders

Hal Shubin, Ajay Divakaran, Kent Wittenburg, Kadir Peker, Regunathan Radhakrishnan

TR2007-118 September 2008

Abstract

We tested our previously reported sports highlights playback for personal video recorders with a carefully chosen set of sports aficionados. Each subject spent about an hour with the content, going through the same basic steps of introduction, trying out the system, and follow up questionnaire. The main conclusion was that the users unanimously liked the functionality very much even when it made mistakes. Furthermore, the users felt that if the user interface were made much more responsive so as to quickly compensate for false alarms and misses, the functionality would be vastly enhanced. The ability to choose summaries of any desired length turned out to be the main attraction.

SPIE, January 2007

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Assessment of end-user response to sports highlights extraction for personal video recorders

Hal Shubin, Ajay Divakaran, Kent Wittenburg, Kadir A. Peker, Regunathan Radhakrishnan
Mitsubishi Electric Research Laboratories, Cambridge, MA 02139

ABSTRACT

We tested our previously reported sports highlights playback for personal video recorders with a carefully chosen set of sports aficionados. Each subject spent about an hour with the content, going through the same basic steps of introduction, trying out the system, and follow up questionnaire. The main conclusion was that the users unanimously liked the functionality very much even when it made mistakes. Furthermore, the users felt that if the user interface were made much more responsive so as to quickly compensate for false alarms and misses, the functionality would be vastly enhanced. The ability to choose summaries of any desired length turned out to be the main attraction.

Keywords: Video Summarization, Subjective Assessment of Summarization

1. INTRODUCTION

Assessment of the success of video summarization has been the next logical step in research on video summarization. In our previous paper [1], we considered success criteria for end-user satisfaction. We framed video summarization as providing a segmentation of content that enables tasks such as getting a digest, skipping over certain segments etc. We then described the role of the presentation-interface and covered some guidelines for the use of video summarization, and described a plan for evaluation.

In this paper, we describe a user study that evaluates the success of a sports highlight playback system. We devised a user interface and asked a group of sports aficionados to evaluate the system. We found that while they made suggestions for improvement, they enjoyed the experience of watching sports highlights even when the system worked imperfectly.

The rest of this paper is organized as follows. In section 2, we briefly describe our past work on sports highlights extraction and describe the video browsing interface. In section 3 we describe the study setup. In section 4 we consider the response from the subjects. In section 5 summarize our results and consider future research.

2. PROPOSED INTERFACE

In our past work [3], we developed a sports highlights extraction scheme based on detection of a mixture of excited speech and cheering. We compute the percentage of the aforementioned mixture in a window centered at the time of interest and assign it as the “interest value.” The output of the summarization is therefore an “interest profile” or “action map” that indicates which portions of the content are relatively more interesting. As illustrated in Figure 1, such an action map enables the user to get highlights of the content by retaining only the parts that exceed a certain threshold of interest. Furthermore, the user can adjust the length of the summary by moving the threshold up or down.

We illustrate our proposed interface in Figure 2. It displays the action map to the user and provides him with the ability to move the threshold up or down. It also provides the user with the length of the resulting summary so as to help him reach the desired length. It also uses a color gradient for the color map so that the color of the points is a different shade depending on the interest value, which allows the user to quickly identify the relative interest values over the time line of the content. It also gives the users the ability to skip ahead or back to the next or previous highlight. We have implemented the interface using Flash. The action maps were computed off-line using our previously proposed technique described above. The success of the interface clearly depends on both the accuracy of the action map and the effectiveness of the interface. Our study targeted the overall success of the system.

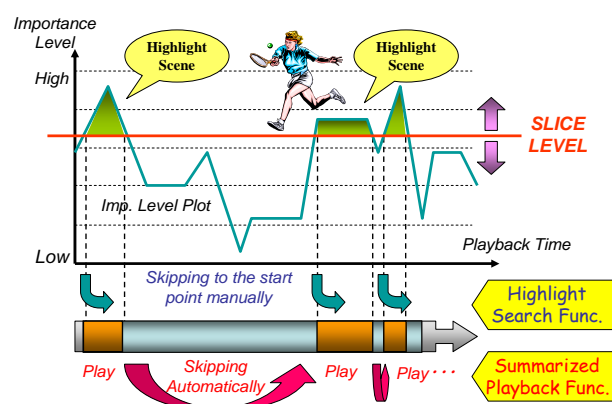


Figure 1: Highlight Scene Detection and Playback



Figure 2: Proposed Sports Highlight Playback Interface

3. STUDY SETUP

One of us sat with the subjects, watching a large-screen TV. Each session included an initial questionnaire on background information, a series of tasks and then asking follow-up questions. Participants used a remote control that was set up for this study. The other authors observed remotely through a video hookup. Note that for the purposes of this paper we use the terms “subject,” “end-user,” “user,” and “participant” interchangeably to mean the sports aficionados who tested our system.

We started each session by explaining how the feature worked, identifying the working buttons on the remote and giving a demonstration. The remote had the usual Play, Fast-Forward (FF) and Rewind (REW) buttons in addition to which it also had the highlight skip forward and back and the Auto-skip buttons. Each participant explored the software by watching one of the Super Bowl recordings. We explained the features again during this exploration and as they did the tasks. They all learned the system quickly

There were three simple tasks, which we described orally. The first two asked participants to use Highlight Viewer to get an overview of a game they'd recorded, assuming that they had (1) ten minutes before they had to leave the house in the morning, or (2) thirty minutes free on a weekend day. In describing the third task, we told them that they had a full afternoon with no expected interruptions and could spend all the time they wanted to watch a program. In each task, participants selected which programs to watch.

To summarize our recruiting procedure, we asked that five or more be male (they all were), the minimum age was 21, all be passionately watching sports programs on TV, recording them digitally to watch later. There were restrictions about the specific sports to ensure variety. Our criteria were driven by the desire to balance the competing requirements of variety and consistency.

Note that our study was not set up as a rigorous scientific experiment, since that would take much more resources than we could afford. The intention was to find out how much our subjects liked the sports highlights playback feature. Note that this is related to but different from rigorously testing how well the highlights extraction carried out its task of extracting highlights.

We administered an initial questionnaire at the beginning of each subject session and ended with a final questionnaire. We have provided the questionnaires in the appendix, which also includes the user responses and the sports content that each subject watched. Salient questions from the questionnaires included:

1. How much of sports video do you record every week/month?
2. What was your overall impression of the interface?
3. Do you think it captured all the highlights?

Note that we attempted to get a profile of the subject as well as his subjective response to the system. Note again that the emphasis was on how much the user liked the system and not on how successful the subjects in carrying out target tasks. The reason for such a focus was that it allowed the experiment to be simple and manageable. Detailed task assessment would have required a much more rigorous and resource consuming procedure.

4. END-USER RESPONSE

We both administered questionnaires as well as recorded comments by the users, since we felt that the numerical data from the questionnaires would not completely capture the end-users' experience. We list the detailed user response to the questionnaires in the appendix. We present a summary of their response here.

Highlights Extraction

What is a highlight? That's an important question to consider as we continue development. P1 and P5 discussed them as "game-changing events" and "plays that changed the game."

Participants liked the system when it found highlights correctly, although some highlights weren't exciting. The meanings of "exciting" and "highlight" are subjective: they depend on whether the action is offensive or defensive (offensive ones being more important); if it's caused by the home team, the away team or a neutral team (e.g., during a Super Bowl, which is generally played in neither team's city); and who the announcers are.

There was a lot of discussion about highlights. Here is a sampling:

- > P1 thought that some highlights seemed "random."
- > P3 said that crowds react to things that he does not always care about.
- > P4 "didn't like" false and missing highlights. He said, "show me home runs and touchdowns."
- > P5 pointed out that people in the crowd "react to disappointments and players they don't like," but that's not useful to him. He gave this example: If Barry Bonds came up to bat with three men on

base and struck out, it's not a highlight because it wouldn't change the game. (In fact, the strikeout did *prevent* a change from occurring). It's "what puts points on the board" that interests him.



- > P6 wanted "a better selection of highlights."
- > P8 wasn't not sure how "they pick highlights" because I didn't tell him anything about it. He said that "someone must sort of arbitrarily pick out where they think the highlights are."

While participants were disappointed at missed or false highlights and mentioned them while rating the system, no one wanted to stop using it because of problems with highlights. The telling question is whether Highlight Viewer (our interface) can get the highlights right. Missed highlights were more of a concern than false highlights (which participants explained away in different ways) or incomplete highlights (which they could recover from). We don't know how long people would tolerate problems with identifying highlights.

Users were mostly able to recover from false or incomplete highlights using the rewind and fast-forward buttons but wanted the system to be more responsive and to provide more past and future context.

The Action Map

The users unanimously liked the action map since it both allowed them to anticipate events of interest and flexibly adjust the length of the summary by choosing a threshold. They wanted the graph to be less obtrusive since currently it obscures a substantial part of the screen.

Moving through a program: Auto skip, (Highlights skip   FF and Rew

The users liked the Auto-skip feature that just plays the portions that are above the threshold since it allows them to sit back after choosing the summary time. They liked the highlights skip and wanted the FF and REW to be more responsive.

Suitability to different sports

The subjects felt overall that soccer, hockey and baseball were better suited for highlight playback. We think that is so because these sports are either deliberate (baseball) or have a small number of scoring events (soccer).

Game Details in the Display

The users were unanimous in thinking that incorporating game details such as scores, averages, player profiles etc. would vastly enhance the enjoyment of the game. Even though this lies outside the scope of automatic sports highlights it does significantly bear on the user enjoyment.

.

5. CONCLUSIONS AND FUTURE WORK

We tested our proposed highlights browsing interface by having eight sports aficionados run the highlight playback using a variety of content. We assessed their response through questionnaires and verbal comments. Since there was a general consistency in the response, we are inclined to believe that the results are reliable even though the sample is small. In short, the users really enjoyed using the system but would like the highlights extraction to be more accurate, the movement through the content to be smoother and more information about the game to be displayed. These results validate our hypothesis in our previous paper [3] that a good interface combined with fair highlights detection accuracy would do better than would a combination of a fair interface with high detection accuracy.

In future work we will work on increasing the accuracy of the highlights extraction as well as on making the interface much more responsive so as to recover from false or partial highlights quickly. Assessment of the user response to the improved system will be the next step.

REFERENCES

1. C. Forlines, K. Peker, and A. Divakaran, "Subjective Assessment of Consumer Video Summarization," SPIE Multimedia Content Analysis, Management and Retrieval, San Jose 2006.
2. L. Xing, H. Yu, Q. Huang, Q. Ye and A. Divakaran, "Subjective Evaluation Criterion for Selecting Affective Features and Modeling Highlights," SPIE Multimedia Content Analysis, Management and Retrieval, San Jose 2006.
3. I. Otsuka, A. Divakaran, K. Nakane, K. Hatanaka, M. Ogawa, "A Highlight Scene Detection and Video Summarization System using Audio Features for a Personal Video Recorder", IEEE Transactions on Consumer Electronics, 2005.

APPENDIX

1. Initial questionnaire

- > Based on the description you read, what do you expect this feature to be like?

- > How do you rate your ability to use computers and other technology?
Example: expert, average, novice.

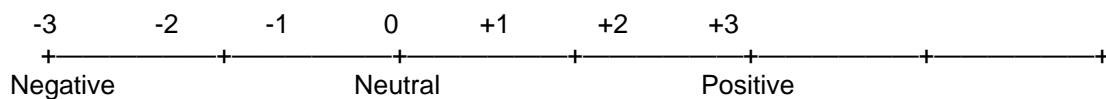
- > What *three* sports do you watch the most? Mark “1”, “2” or “3” next to the top three.

<input type="checkbox"/> Football	<input type="checkbox"/> Baseball
<input type="checkbox"/> Basketball	<input type="checkbox"/> Hockey
<input type="checkbox"/> Soccer	<input type="checkbox"/> Golf
<input type="checkbox"/> Tennis	<input type="checkbox"/> Other: _____

- > How many sports programs do you record in a week or month?
Which ones typically?

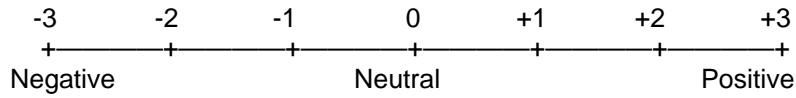
- > Do you wind up watching them? Yes / No
When?
How do you watch them (start-to-finish, skipping, other)?

- > **[Start the first task. After 5 minutes, get an initial rating of the UI.]**



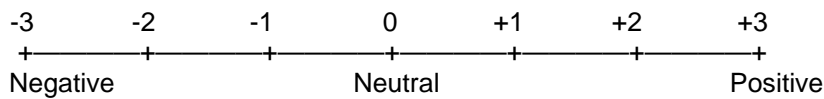
2. Final questionnaire

- > You just spent some time with the highlights viewer. What is your overall impression of it?



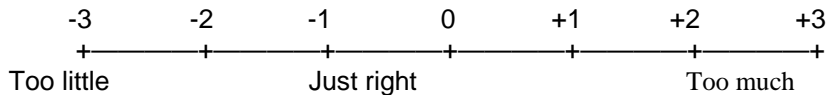
- > [Restate initial expectations] Now that you've used it, how does the experience compare with what you expected?

- > How did it compare with the way you watch recorded sports programs now?

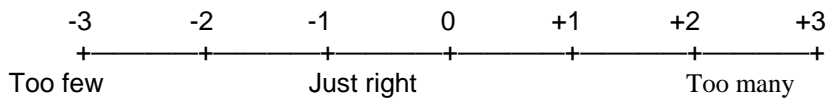


- > Would this feature let you do any more or less than you do now?

- > Think about a particular highlight that you just watched. (Maybe look at it again.)
Did it show you the right amount of information?



- > Did the system get all the highlights?
Did it highlight anything that it should *not* have highlighted?



- > We looked at a few different ways of working. What did you think of them?

Note: *We don't have these implemented, but we can discuss them.*

"Autopilot" –

"Semi-active" –

"Manual" –

- Did this feature seem more appropriate for one over the other?

-
- 3 -2 -1 0 +1 +2 +3
- Unclear Neutral Clear

-
- A horizontal number line with tick marks at -3, -2, -1, 0, +1, +2, and +3. Below the line, the word "Negative" is written under -3 and -2, "Neutral" is written under 0, and "Positive" is written under +1, +2, and +3.

-
- 3 -2 -1 0 +1 +2 +3
- Too little Neutral Too much

- > Any final comments?

Demographic information				Programs the participant had seen										Program chosen plus initial threshold level				
#	Recordings per month	Favorites to watch	Age	World Series	NCAA	Super Bowl	Senior Bowl	British Open	NHL game	World Cup	Semi-finals	Finals	Program	Level	Program	Level	Program	Level
P1	4+	Baseball Football Basketball	23	✓		✓	✓			✓	✓	✓	World Cup	9	Senior Bowl	6	World Series	
P2	2 to 3	Football Hockey Baseball	46	✓		✓		✓					World Cup	9	British Open	?	World Cup World Series	
P3	4+	Football Basketball Baseball	48	✓		✓	✓	✓		✓			World Series	6	British Open	2	Hockey Football	
P4	2 to 3	Football Basketball Soccer	30	✓		✓	✓			✓			Super Bowl	8	World Cup	4	World Series	
P5	4+	Football Baseball Basketball	33	✓	✓	✓	✓			✓			World Series	8	World Cup	7	Senior Bowl	
P6	4+	Baseball Football Golf	49	✓		✓		✓					World Series	7	British Open	5	NCAA	
P7	2 to 3	Hockey Football Soccer	25			✓				✓			NHL	8	World Cup	5	Senior Bowl	
P8	4+	Football Basketball Hockey	58	✓		✓		✓				✓	World Series	8	British Open	3	Australian Open	

Table 1: List of Sports Content watched by each Subject

#	Question	P1	P2	P3	P4	P5	P6	P7
1	How do you rate your ability to use computers and other technology?	+2	+1.5	0	+2	+1	+1	+2.5
2	What is your initial reaction to the Highlight Viewer after a few minutes using it?	+1	-1.5	+1	+1	-1	+1	+1.5
3	You just spent some time with the highlights viewer. What is your overall impression ?	+1	-1	+3	+2	0	+3	+2
4	How did it compare with the way you watch recorded sports programs now?	+1	0	+3	0	+2	+3	+2
5	Think about a particular highlight that you just watched. Did it show you the right amount of information?	0	-1	0	-1.5	-1.5	+1	-1
6	Did the system get all the highlights?	+0	-2	0	-1	-1.5	0	-0.5
7	How clear was it to use the graph showing highlights and the remote to move through the video?	+2	+3	+3	+3	-0.5	+3	+2.5
8	What did you think of the transition when the video moved from one highlight to another?	+1	+1	+2	+2	--	-1	0
9	Was there enough feedback? Did you know what you could do at any time? Did you know what the system was doing and whether it was doing what you asked?	+1	+1.5	0	0	-1	+2	0

Table 2: Subjects' Response to Questionnaire