

ViSA: Video Segmentation and Annotation

Afzal Godil

National Institute of Standards & Technology

afzal.godil@nist.gov

<http://www.nist.gov/webmet>

ABSTRACT

Screen video recording is a common component of usability testing. We have developed tools for reviewing and analyzing the video more efficiently. The tool: 1) breaks up a video into shorter segments and provide a compact pictorial summarization of the video; 2) provides a variety of ways for accessing video; 3) allows web-based review of video; and 4) provides a web-based way to post and view annotations of the video.

INTRODUCTION

Observation of a user is the foundation of usability testing. Traditionally usability experts have directly observed test users as they interact with software or web sites. Remote observation based on video screen capture and audio capture offers an alternative [1] to direct observation. This has created an impetus for creating tools to help analyze the video more efficiently. We describe a web-based review and summarization system of the screen capture video, called ViSA (Video Segmentation and Annotation). Major benefits of the web-based system include the ability to allow usability experts to review the video from any computer on the intranet. Automatic summarization based on video segmentation is one such method [2]. It creates a subset of keyframes which contain almost as much information as the original video. From the summaries provided the usability expert can quickly find parts of the video which are interesting. ViSA can also index the video based on audio signal graph and the transcript. It will also play the video from that point in time by clicking on the graph or transcript. Finally it is possible to post and view annotations of the screen video and to create anchor points into the video. The topic of annotating videos has been addressed previously by Barger et al [3]. The tools we have developed could work equally well for any video source.

DESCRIPTION OF VIDEO TOOLS

Video Screen Capture

We used Camtasia¹ (TechSmith Corporation) to capture screen videos (Figure 1) and selected the option for recording audio; AVI files were created. Reviewing the video allows one to measure a variety of things, such as, how long it takes a user to figure out a navigation scheme or how many times they click on the wrong link. However, by using interaction and audio recording software, details can emerge that end up being the most important finds -- such as mouse hesitation, gestures of irritation, and sound of confusion in the user's voice.

Video Segmentation Tool

The video segmentation tool, shown in Figures 2 and 3, reads in the captured AVI video, finds the keyframes based on the difference in the color histogram between two successive frames and creates a filmstrip of keyframe images. The generated keyframes are a compact pictorial summarization of the usability testing video. The keyframes images are also hyperlinked back to the video -- clicking the keyframe starts the video playing from that point in time. The code also embeds the images of the keyframes into an html page and provides links to the captured video using JavaScript functions.

¹ Mention of trade names does not imply endorsement by NIST

Audio Signal Graph and Transcript

The audio track can be used to create an audio signal graph (Figure 4). The trace shows times at which the user was speaking or shouting, etc. The audio graph is also hyperlinked to the video so that clicking on the graph causes the video to play at the selected time point. The audio can also be used to create a transcript of voice of the speaker (Figure 5), which is also hyperlinked to the video. In our case this was done by hand, but could be transcribed automatically using speech recognition.

Web-Based Review of Video

The interface for web-based video review is shown in Figure 3. The panel on the left side shows all the videos that are available for review and also has a simple search capability. There are three areas on the right side. The top-most is the control frame and is used to select the type of retrieval -- storyboard, transcript, or audio signal graph. The area below that embeds the Real media plug-in to show, play, pause and stop the video. Below this there are two buttons to post and view annotations. The bottom-most part of the main area shows the compact pictorial summarization/thumbnail of the usability video as a filmstrip of keyframes images, transcripts or audio signal graph.

Annotation of the video

We also have created a web-based way for the reviewer of the video to post comments or annotate the video to indicate "interesting things" happening in the video as shown Figure 6. Annotations can be reviewed later and can be used to cue the video by clicking on the movie icon (Figure 7). This web-based way to post and view annotations is based on CGI and forms and is written in Perl.

SUMMARY

ViSA is a tool that can be useful for a usability expert for reviewing and analyzing of screen video recording. The tool can perform segmentation of video to create a filmstrip of keyframes for summarization. From the summaries the usability expert can quickly review parts of video which are interesting. It also allows web-based review and a variety of ways to accessing the video. Finally, it provides a way to annotate the video for posting notes for review later.

REFERENCES

1. S. Thompson, Remote Observation Strategies for Usability Testing, http://www.lita.org/Content/NavigationMenu/LITA/LITA_Publications4/ITAL__Information_Technology_and_Libraries/2201_thompson.htm
2. P. Chiu, A. Girgensohn, W. Polak, E. Rieffel, and L. Wilcox. A Genetic Algorithm for Video Segmentation and Summarization .In Proceedings of IEEE International Conference on Multimedia and Expo, vol. III, pp. 1329-1332, 2000.
3. D. Barger, A. Gupta, E. Sanocki and J. Grudin. Annotations for streaming video on the web: System Design and Usage Studies. In Proceeding of WWW8, pp. 61-75, 1999.

ABOUT THE AUTHOR

Afzal Godil is a Computer Specialist in the Visualization and Usability Group at NIST where his duties involve development of tools and techniques in area of 3D graphics/visualization, computational methods and pattern recognition He has an MS in Aerospace and Mechanical Engineering from the University of Arizona.

Figure 1. Camtasia Video Screen Capture

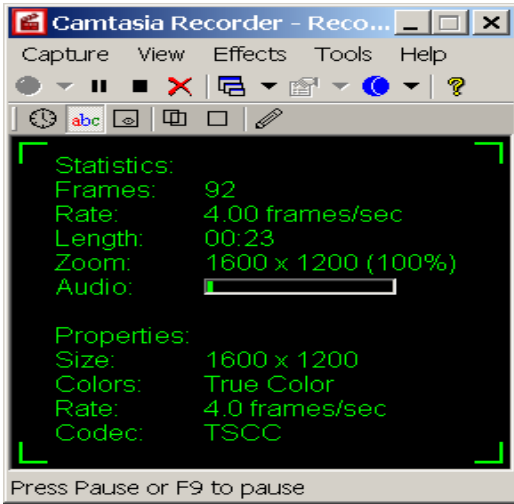


Figure 2. Video Segmentation Tool

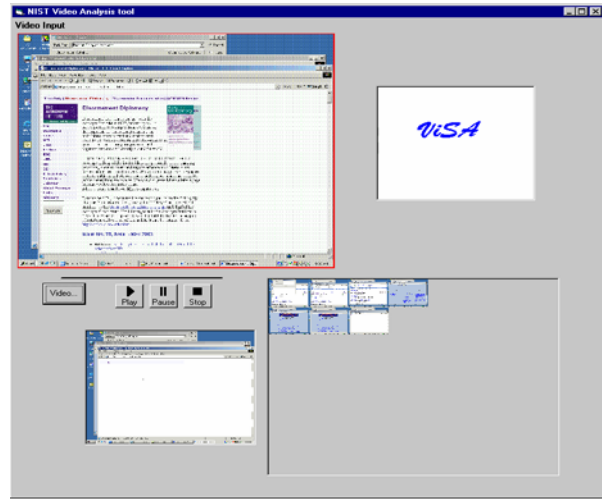


Figure 3. Web-Based Review of the Video using thumbnails produced during segmentation

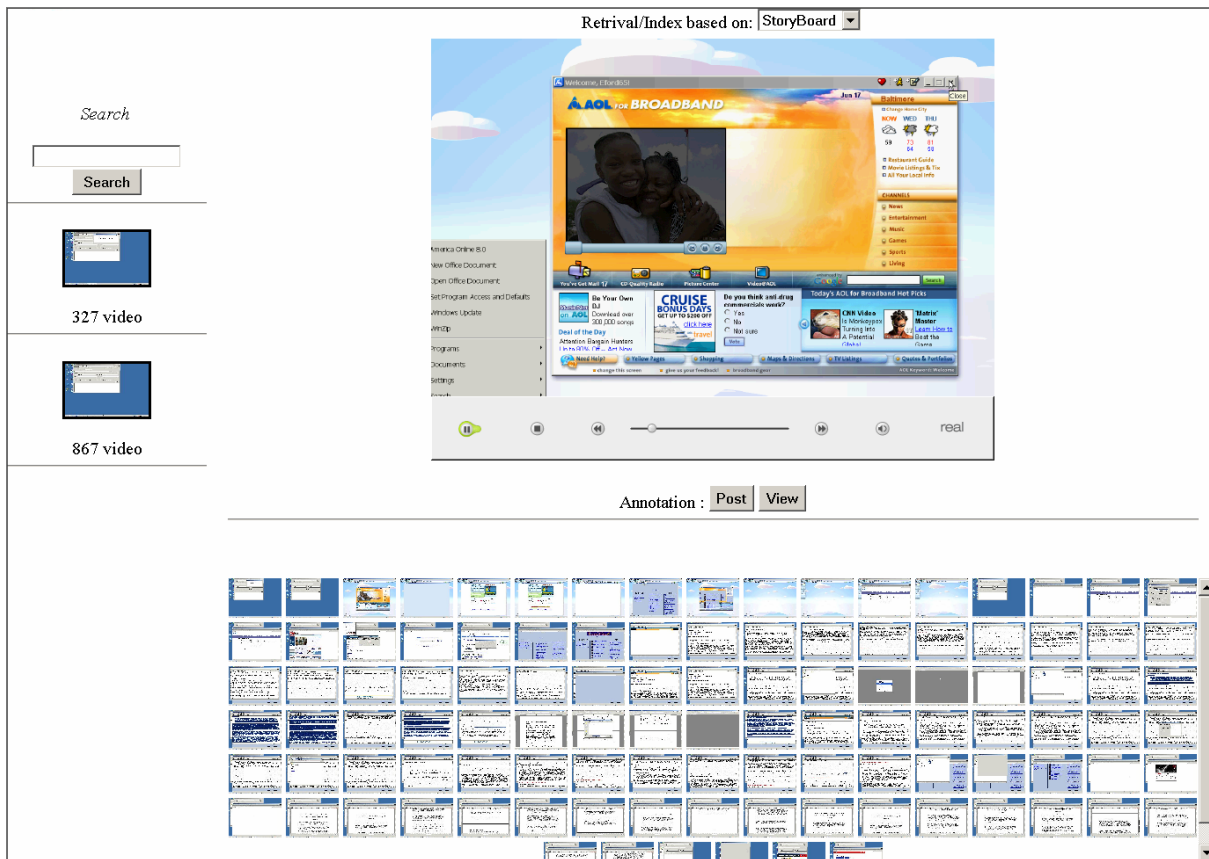


Figure 4. Transcript based retrieval

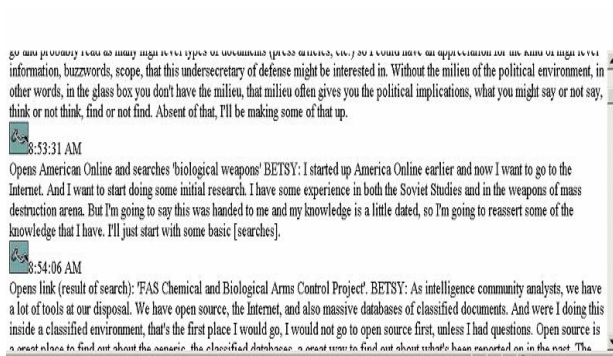


Figure 5. Accessing the video from an audio signal graph

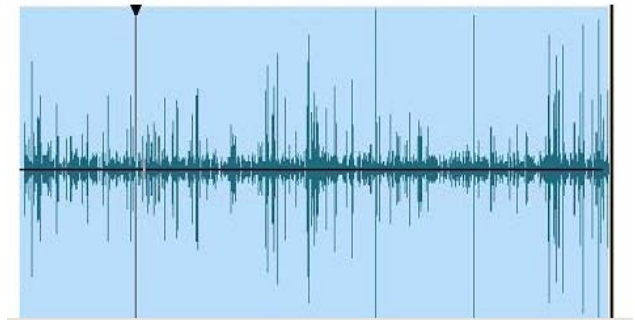


Figure 6. Post an annotation over the Internet



Figure 7. Viewing an annotation

The screenshot shows a web browser window displaying a table of annotations. The table has the following columns: Link, Name, Time, Comments, and database. The data rows are as follows:

Link	Name	Time	Comments	database
	godil	97600	test2	Query
	jean	909416	test44	Query
	godil	383177	nice stuff	Query
	joe	99600	test1	Query
		167840		Query
	vvv	0	qqq	Query
	jjj	3100	hhh	Query
	john	127330	Aol started	Query
	john	127330	Aol started	Query
	Public		audio	Query
	mike	712018	good job	Query