

# Enriched Video Semantic Metadata: Authorization, Integration, and Presentation

## Xiangming Mu

Interaction Design Lab, School of Information and Library Science  
University of North Carolina at Chapel Hill, Chapel Hill, NC27599-3360  
Phone: (919) 962-8366 Email: mux@ils.unc.edu

## Gary Marchionini

Interaction Design Lab, School of Information and Library Science  
University of North Carolina at Chapel Hill, Chapel Hill, NC27599-3360  
Phone: (919) 962-8366 Email: mux@ils.unc.edu

### ABSTRACT

**An enriched video metadata framework including video authorization using VAST, metadata integration, and user level applications is presented. The Video Annotation and Summarization Tool (VAST) is a novel video metadata authorization system that integrates both semantic and visual metadata. Balance between accuracy and efficiency is achieved by adopting a semi-automatic authorization approach. Semantic information such as video annotation is integrated with the visual information under a XML schema. Results from user studies and field experiments using the VAST metadata demonstrated that the enriched metadata were seamlessly incorporated into application level programs such as the Interactive Shared Educational Environment (ISEE). Given the VAST metadata, a new video surrogate called the smartlink video surrogate was proposed and deployed in the ISEE. VAST has become a key component of the emerging Open Video Digital Library toolkit.**

### Introduction

The development of digital libraries and “open” video projects such as the Internet Archive Project (<http://www.archive.org>) and the Open Video Project (<http://www.open-video.org>) provides an increasing volume of digital video documents on the Internet. Compared to analog video such as tapes, digital video is not only much easier to store and to transmit, but also provides a means for browsing the video content quickly instead of taking a longer time to view the full length video.

Video surrogate (Ding et al., 1999) provides a concise representation of the video while preserving the essential messages (Komlodi & Marchionini, 1998). Video surrogate is also referred to as video summary (Yeo & Yeung, 1997), video abstraction or video abstract (Lienhart et al., 1997) and can be classified into textual surrogate and visual surrogate. Metadata such as title, publisher, date, content abstraction, closed-caption data, and/or full-text transcript, are often utilized in textual video surrogates. Visual surrogates usually refer to video frames or a ‘skimmed’ video of the original (Christel et al, 1999). In some cases one poster frame is used to represent a video clip, while in other cases a set of images is displayed as a filmstrip or storyboard, in which a series of thumbnails, also called key frames, are presented, each representative of a shot or a scene of the video. By dropping frames, skimmed video enables users to view the entire video in a much shorter time—a type of fast-forward. One challenge in video metadata authorization is how to integrate the textual video metadata with the visual video metadata to provide an enriched video surrogate.

Video surrogates can be authorized by two means: by humans or automatically. Manual authorization is usually very time-consuming but more accurate (He, et al, 1999). This might be one of the reasons why there are so few video surrogates in general use. Automatic metadata authorization is usually based on the video’s physical features such as color, motion, shape, or brightness data. How to balance the tradeoff between quality and efficiency in the video metadata authorization process is another challenge.

The Video Annotation and Summary Tool (VAST) is a semi-automatic two-phase video metadata authorization environment that seamlessly incorporates textual video

metadata such as annotation with related visual metadata, currently including key frames and fast-forward surrogates. By automatically extracting every Nth frame in sequence, a reduced number of frames are cached, which is referred to as the primitive video frames. In the second phase, key frames that represent the essential information of the original video are manually selected. Annotations may be linked to the related key frames at this phase. At the indexer's option, fast-forward surrogates can be directly formed based on the phase one primary frames.

Video's temporality distinguishes this format from other media. Frames are distributed along a timeline and each frame has a unique timestamp that represents its temporal position. VAST identifies the associated timestamp of each frame and integrates it with other metadata. Finally, the metadata is encoded in a XML schema, which can be adapted directly to various application level programs. Thus, VAST serves as a key component in our digital video indexing.

This paper is organized as follows. After a brief introduction to related research, the user interface and system architecture of VAST are presented. In the next section, the video XML metadata schema is introduced followed by an example of the application of the metadata in a multimedia distance learning tool. Next, two user studies using the fast-forward video surrogates produced by VAST and two field studies using the distance learning tool that adopted the VAST key frames and annotations are summarized. Finally, a novel video surrogate that combines both text transcripts and video links is introduced

## Related Work

Video surrogates are routinely adopted in the user interfaces of many digital library systems to facilitate content browsing; examples of the use of such surrogates include the Open Video Project from University of North Carolina at Chapel Hill, the Informedia Project from Carnegie Mellon University, and cueVideo from IBM. The most widely used video surrogate is the video storyboard.

Video visual structural characteristics such as color, texture, and shape can be utilized to detect shot boundaries and to select key frames (Aigrain et al, 1996). Various algorithms have been developed to automate such video partitions and key frame selections. In general these algorithms can be roughly classified into two main approaches: one is based on the statistically significant differences of changes and the other is based on the modeling of motion or image content (Lindley, 1997). Additionally, camera operations (motion, pan/tilt, and zoom) are also helpful because they implicitly express the intentions of the video producers and thus can be used to aid the scene detection (Aigrain et al, 1996). The biggest attractions for the automatic techniques are efficiency and scalability.

However, as argued by Lindley (1997), automatically generated visual description "alone provide very limited effectiveness for applications concerned with what a video stream is 'about'". Thus, the rich text that presents the semantic perspective of the video is still necessary, especially in educational domains where the "talking head" lectures provide limited visual information and the transcripts are usually critical.

Srinivasan et al. (1997) argued that "effective browsing through video material requires a combination of the use of the actual video material and the various other representations of the material". Two object models were defined by Srinivasan: structural objects and semantic objects. The structural objects represent structural components of the video such as frames, camera motions and shots. The semantic objects model represents the concepts associated with the video, including catalog descriptions of the content of the video, segment descriptions, textual dialog transcripts and shot lists. Extended by Lindley (1997), four levels of video semantics were proposed: the diegetic level, which designates the sum of the video's denotation; the connotative level, which designates the metaphorical, analogical and associative meaning of the video; the subtextual level, which corresponds to the hidden and suppressed meanings; and the cinematic level, which is concerned with the expressive artifacts. A study from the Informedia project at CMU has shown that text-augmented storyboards performed significantly better than storyboards with no text (Christel & Warmack, 2001). After comparing different treatments including images only, image with text, images with text but in different layout and length, Christel et al. (2001) indicated that participants preferred the storyboards with text and achieved faster task times. The results are in agreement with Ding's finding that images with text drew more participants' attention (Ding et al, 1999).

Although semantic information can be generated automatically from phrase partitioning, captioning, and capitalization processes (Christel & Warmack, 2001), other scholars argue (Aigrain et al., 1996) that in a documentary, "there is no one-to-one correspondence between the video track and an audio track". The transcripts extracted are not synchronized perfectly with the video. Realizing this inconsistency, Myers et al. (2001) proposed a means to fill in the gaps between different lengths of video and audio. However, higher quality transcripts are still generated more by humans than by automation. A user study demonstrated that automatically generated video summaries were less clear, concise, and coherent (He et al., 1999). More critically, manual annotation can provide customized semantic information and thus enrich the automatically generated metadata. Some video authorization tools have been developed to facilitate this type of metadata generation.

The A4SM (Authoring System for Syntactic, Semantic and Semiotic Modeling) project (Nack & Putz, 2001) proposed a semi-automated annotation system that was able to build structured annotations on the fly. The annotation was organized at the segment level and was incorporated into an XML schema. However, the application domain of the tool focused on the video genre of news and provided no support for frame level annotation operations. VideoLogger from Virage™ (<http://www.virage.com>) enables users to generate frame thumbnails on shot level. A sensitivity control mechanism is also supported; however, users cannot change the size of the thumbnails. Timestamps are automatically added to each thumbnail but no direct annotation functions are supported on the frame level. Even though the thumbnails can be published into storyboards in HTML format, it is not convenient for further manipulation on the data from the program developer's perspective.

The VAST (Video Annotation and Summary Tool) described here aims to overcome some of these limitations in existing work and is distinguished from other video metadata authorization tools in the following ways:

1. Supports semi-automatic video metadata authorization.
2. Supports user controlled key frame generation.
3. Timestamps are integrated into frame level metadata.

XML schema is adopted to provide flexibility and reusability

### VAST: User Interface

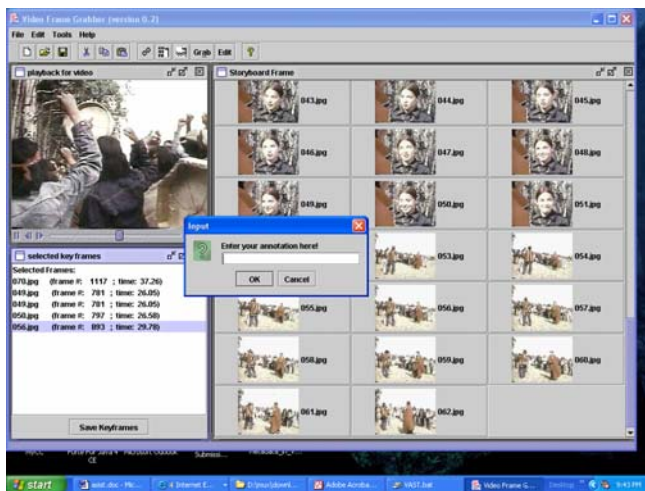


Figure 1: the VAST user interface

VAST provides a desktop style Graphic User Interface (GUI) that supports both semantic information authoring function and manual key frame selection function. Figure 1 is a sample user interface of the VAST. At the top of the desktop is the control panel. Beneath it on the right is the

storyboard panel where frames generated from the video can be displayed. To the top left is a video player which enables users to monitor the playback of the video. A slide bar beneath it provides VCR video navigation support. Under the video player on the bottom left is the video annotation panel. A popup window appears for the user to add text annotation for a selected frame.

Flexibility is provided for users of the interface by allowing them to freely rearrange the layout of the GUI in accordance to their personal preferences. The positions of various components can be dragged and dropped on an arbitrary position within the desktop. The size of each component is also changeable. For example, the user can enlarge the video player and place it in the center of the interface while watching the video, and minimize or close it while editing the metadata.

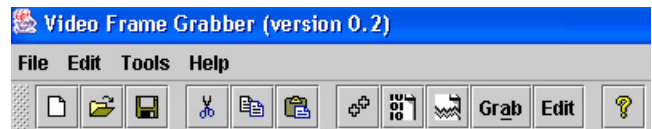


Figure 2: The VAST control panel

Figure 2 illustrates the control panel that contains a set of control buttons. In addition to the traditional file operation buttons (for opening/closing files), metadata manipulation buttons are added to support quick manipulations. The button with two cross signs is the configuration button and is used to open the configuration dialog window. Figure 3 illustrates the dialog window that controls the extraction of the primary frames, which are not the key frames of the video, but a pool of key frame candidates automatically generated from the video by selecting every Nth frame. Currently two control variables are supported: the image grabbing rate variable, which controls how many frames are dropped before a primary frame is selected (in other word, the value of the N), and the image size variable, which controls the size of the primary frame. The default value for the image grabbing rate is sixteen, which means that fifteen frames will be dropped before a primary frame is grabbed. "0.3" is the default setting value for the image size, which means the primary frame will be a thumbnail with size 0.3 times the original frame size. By adjusting this number, the user can get small size thumbnails or large size poster frames.

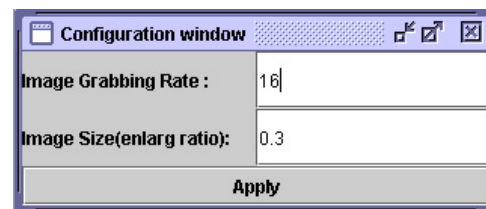


Figure 3: VAST configuration panel

Next to the configuration button is the primary frame displaying button. Clicking on this button will update the

frames displayed on the storyboard panel (if the frames have been updated). Only the latest generated 2000 primary frames are available for display in a session. All the generated primary frames are automatically saved to user's local storage. Users can manually select the key frames from the displayed primary frames by a single click on the target. The selected item will automatically be presented on the annotation panel on the bottom left of the screen. The timestamp and frame number are automatically displayed with each selected frame item (Figure 1). Double clicking on any item listed in the panel will popup a dialogue box that allows users to add/edit the annotation which is associated with the frame.

The next button is a "clear panel" button whose function is to clear all the images displayed in the storyboard panel. The button with the text "grab" is used to grab a frame on the fly and add it as a key frame while the user is watching the video, thus allowing manual selection of frames by an indexer. The "edit" button is to add/edit/delete an annotation for any selected frame in the annotation panel using the dialogue box (Figure 1). Help information is available by pressing the question mark button.

The video player is used to view and monitor the video playback. The default position for the video player is on the top left. Implemented in Java and the Application Program Interfaces (APIs) from Java Media Framework (JMF), this video player supports most of the current video formats (<http://java.sun.com/products/java-media/jmf/2.1.1/formats.html>) including AVI, MPEG, and QuickTime. A visual VCR like control component just beneath the video player provides users manipulation functions such as pausing/resuming. A slider is also provided as default which enables users to jump to an arbitrary position along the video timeline. A tooltip, which indicates the current timestamp of the video, appears automatically when the user's mouse is over the slider for several seconds.

A fast-forward version of the original video can be produced by presenting primary frames with the normal speed. With the frames available in sequence, it is straightforward to generate the fast-forward video with some commercial tools such as QuickTime ([www.apple.com/quicktime](http://www.apple.com/quicktime)). The time compression ratio of the fast-forward is dependent on the parameter of the *Image grabbing ratio*. For example, choosing the default value of sixteen, the fast-forward will play sixteen times faster than the original one.

Finally, all the frames as well as the associated metadata including the timestamp, frame number, name, and annotations are integrated into a XML schema, which we will introduce next, and is stored in the user's local computer as a file document.

## VAST: Metadata and XML Schema

There are two means to organize video metadata: hierarchical and sequential. *"In the hierarchical mode, the information is organized into successive levels, each describing the audio-visual content at a different level of detail. In general, the levels closer to the root of the hierarchy provide more coarse summaries, and levels further from the root provide more detailed summaries. The sequential summary provides a sequence of images or video frames"* (MPEG-7,2002). A hierarchical structure is adopted to represent VAST metadata.

To integrate the visual and semantic information we generate from VAST, a preliminary hierarchical XML schema was developed which includes the following levels: the segment level, the scene level, the shot level, and the frame level. Our schema is an extension of the MPEG-7 video navigation and access Descriptor and Description Schemes and encoded in the MPEG-7 DDL.

The sketch of a coded metadata schema in VAST looks like the following code:

```
<program xmlns =
http://www.ils.unc.edu/VAST/ExampleSchema>
  <generalInfo>
    <title>...</title>
    <annotation>...</annotation>
    ...
  </generalInfo>
  <visualInfo>
    <keyFrames datatype = "IntegerVector" size = "16" >
      <kFrame >
        <sequenceNo datatype = "unsigned8">
1 </sequenceNo>
        <annotation> ... </annotation>
        <timestamp datatype = "RelTimePoint"
minValue = "0.0">
        ...
      </timestamp>
      <duration datatype =
"FractionalDuration" > ... </duration>
      ...
    </kFrame>
    ...
  </keyFrames>
  <visualFF >
    <location datatype = "String"> ... </location>
```

```

<startPoint datatype = "RelTimePoint"> 0.0
</startPoint>
<endPoint datatype = "RelTimePoint"> ... </endPoint>
<ffDuration datatype = "RelTimePoint">
..</ffDuration>
<compressionRatio datatype="unsigned8" >
...</compressionRatio>
<generalInfo>
    <author> ..</author>
    ...
</generalInfo>
<format datatype= "String"> QuickTime </format>
...
</visualInfo>
...
</program>

```

## VAST: Applications

VAST was originally created to support creating storyboards and fast forwards for use in the Open Video Digital Library. With the help of VAST, textual metadata was authorized and was further integrated with the visual metadata into the repository database. Over the past two years, VAST has found other uses, including as the basis for customized video surrogates that used for user studies and for a new system that supports remote video study (ISEE), and a forthcoming video indexer's toolkit.

### Open-Video project

The Open-Video Project has more than eighteen thousand digitalized video segments with lengths from several seconds to nearly an hour. Most of the videos have automatically extracted key frames associated with them and VAST has been used to create storyboards for most of these videos to date. At this time, fast forward surrogates have been created for about one-half of the collection.

The VAST was also utilized for authoring fast-forward videos for two user studies conducted in the Interactive Design Lab (IDL) in University of North Carolina at Chapel Hill. In the first study, five surrogates including fast-forwards were evaluated in terms of their usefulness and usability in accomplishing specific tasks (Wildemuth et al., 2002). The compression ratio of the fast-forward was the default value, which means they are sixteen times faster than the original videos. In the second study various fast-forward speeds were studied to try to answer a simple question: how fast is too fast? The results for this study are presented in a different paper (Wildemuth et al, 2003)

### Interactive Shared Educational Environment: ISEE

The Interactive Shared Educational Environment (ISEE) is an advanced video application system that supports highly interactive collaboration distance learning (Mu & Marchionini, 2002). VAST provides both the semantic and visual metadata for the ISEE. A metadata XML parser built in the ISEE accepts and parses the metadata based on the same XML schema as the VAST. Figure 4 illustrates the Graphic User Interface (GUI) of the ISEE. A Video Player, an Interactive Chat Room (ICR), a video Storyboard, and a Shared Web Browser (SWB) are supported in the ISEE based on the VAST metadata. For example, a single click on a specific frame on the storyboard will update the video player to the same time stamp and begin to play there. Timestamp is integrated by VAST as part of the metadata and was loaded by ISEE into the storyboard. The video player modality accepts the signals passed from the storyboard and triggers actions to start, stop, jump back, or jump forward. Another visual metadata, fast-forward video, has also been supported in the new ISEE version although not being tested.

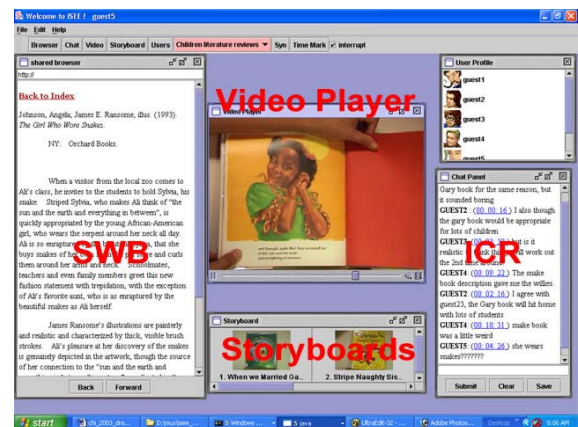


Figure 4: Interactive Shared Educational Environment user interface

### SmartLink in ISEE: a new video surrogate

The SWB not only enables users to share web information, but also provides a platform to display a novel video surrogate: the smartlink surrogate.

Previous studies have indicated the critical role of text in the visual representation of video, like the text in the storyboard (Christel & Warmack, 2001; Ding et al., 1999; Wildemuth et al, 2002). Videos in the educational domain usually have rich semantic information such as the transcript and the annotation. However, the limited screen real estate constrains appending large chunks of text. Research also suggests that separating the text from the images was not a good design approach (Christel & Warmack, 2001) because more effort in terms of eye

movement is required in such a design when users attempt to synchronize the text with the separated images.

By providing video links in a textual document, the smartlink video surrogate is particularly suitable for presenting text-enriched video metadata. Smartlink is based on the integrated metadata produced by VAST. Each link is associated with a segment of the video and both share the same timestamp. A single click on these links will immediately update the video player to the corresponding point based on the shared timestamp (figure 5).

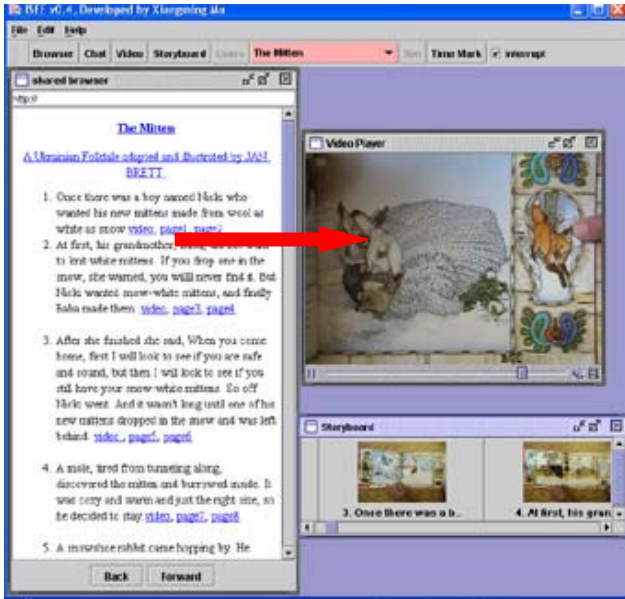


Figure 5: Two types of video surrogates in ISEE: storyboard and smartlink

### Enriched Video Metadata Framework

Figure 6 illustrates the enriched video metadata framework we have discussed. Video visual metadata is integrated with semantic metadata, and then is encoded into an XML schema after authorized in the VAST environment. ISEE is one of the user level applications that adopts the structured video metadata and presents them in various user level video surrogates: video storyboards, video scripts, and the smartlink surrogate.

Even though the VAST can directly feed video metadata to the user applications, the de-coupling of the application from the video metadata authorization provides more flexibility and scalability. It is straightforward for other applications besides the ISEE to utilize the metadata based on the same XML schema. Modifications or the addition of new modalities in the applications do not affect the video metadata authorization.

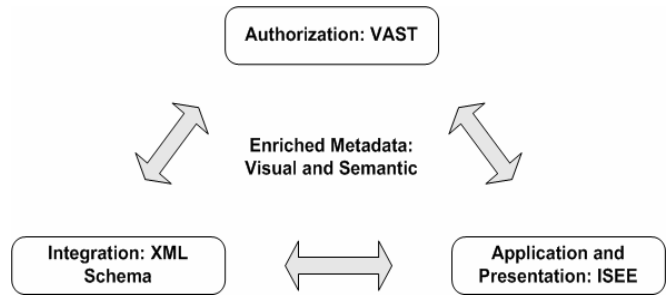


Figure 6: Enriched Video Metadata Framework

### Field Experiments

It is fundamentally difficult to directly evaluate the metadata authorization such as VAST or the metadata schema we developed in the Enriched Video Metadata Framework. Thus user level application tool such as ISEE is utilized to help us understanding the role of video metadata authorization. Two field experiments using ISEE to simulate distance multimedia communication were conducted in November, 2002 and January, 2003 respectively. Both of the studies were in the computer lab of the School of Information and Library Science. Each workstation in the lab has a 10Mb fast Ethernet connection and the ISEE was pre-installed. Each student worked at an individual workstation and wore headphones. Talking was not allowed during the entire study. Thus, the session served as a reasonable simulation for a distance learning setting.

In the first study, volunteers were recruited from a population of students in a Children’s Literature & Related Materials course. Twenty-eight students participated in the study. Participants were asked to assume that they were librarians and should order three children’s books for their library from five available selections. This was a three-phase study. Text preview was available in phase one, then the online chat channel was added in phase two. In phase three both the video and storyboard were added. There were a total of 73 clicks on the video storyboards in phase three, which means that on average each participant clicked the storyboard 2.5 times. The trend of the clicks on the storyboards shows that the number of clicks decreased quickly along the timeline. Participants felt more comfortable using the ISEE system with enriched video metadata ( $t=4.42$ ,  $df=27$ ,  $p<0.001$ ) where video, video storyboard and text reviews were all available. Results also indicated such an enriched environment provided a more effective context for completing the decision-making tasks compared to the phase two (no video and video storyboard) ( $t = 10.5$ ,  $df = 27$ ,  $p<0.001$ ). The details for this study were presented in another paper (Mu et al, 2003)

The second experiment was an informal test which extended from the first study. It was conducted by one of the authors in a class of thirty students to simulate distance-learning class. Twenty-four students attended the class and no one had ever used ISEE before. Based on the same VAST XML video metadata, both the smartlink video surrogate and the video storyboard were available in this session. After a very brief tutorial, students were asked to view a video called "Senses and Sensitivity" which was downloaded from the Open-Video digital library (<http://www.open-video.org>). With both surrogates available, we found that there were a total of 43 clicks on the smartlinks and only 25 clicks on the storyboards, which might indicate that the smartlink was probably also an effective video surrogate. Formal studies are needed to fully evaluate this new video surrogate before we can issue any further conclusions on it.

## Conclusions

Video surrogates such as storyboards enable users to quickly grab the essential information of a video before spending a longer time to view any video media in full length. Key frames that are displayed on the storyboard can be authorized manually or using some automatic technologies. In order to balance the accuracy and efficiency of the two different approaches, a semi-automatic video metadata authorization system called VAST has been developed. The automatic process generates a series of key frame "candidates" (primitive frames) for users to manually select the key frames. The number and the image size of the primitive frames are reconfigurable. The fast-forward version of the original video can also be authorized based on the primitive frames. An XML schema was developed to encode the enriched video metadata which includes both visual and semantic information.

Semantic information, including the video titles, synopses, transcripts, and video annotation is integrated with the visual surrogates in the VAST. The Interactive Shared Educational Environment (ISEE) is a novel multimedia distance learning system developed in the Interactive Design Lab, University of North Carolina at Chapel Hill that uses VAST enriched video metadata for both the storyboard and the smartlink video surrogates. Hyperlinks are made between text and video metadata which enables users to directly access the desired video segment with a single click. Our informal field experiment indicated promising results for this new video surrogate. User studies have been planned in order to get a better understanding about the smartlink video surrogate. The VAST tool itself has become an important component of the Open Video Digital Library toolkit. One direction of continued development is to use it as the basis for a general digital video indexing software suite.

## ACKNOWLEDGMENTS

This work was supported by NSF Grant IIS #0099638. The two fast-forward studies were conducted by the IDL group and one of the field studies was conducted by Amy Pattee and the authors. Dr. Barbara Wildurmth provided help in the study design and preparation of the IRB forms.

## REFERENCES

- Aigrain P., Zhang H., and Petkovic D. 1996 "Content-based representation and retrieval of visual media: A state-of-the-art review", *Multimedia tools and applications* 3,179-202, Kluwer Academic Publishers, The Netherlands
- Christel, M.G., Hauptmann,A.G., Warmack, A.S. & Crosby,S.A. (1999). Adjustable filmstrips and skims as abstractions for a digital video library. In *Proceedings of IEEE forum on Research and Technology Advances in Digital Libraries ADL'99*, 19-21 May 1999, Baltimore, Maryland.
- Christel,M., and Warmack,A.S. The effect of text in storyboards for video navigation. *Proc.IEEE international conference on acoustics, speech, and signal processing(ICASSP)* (Salt Lake City, UT, May 2001). Vol. III. Pp.1409-1412.
- Ding,W., Marchionini,G., & Soergel,D. (1999). Multimodal surrogates for video browsing. *DL'99, Proceedings of the Fourth ACM conference on Digital Libraries*, Berkeley, CA, August 11-14,1999.
- He,L., Sanocki,E., gupta,A., and Grudin,J., 1999, Auto-Summarization of Audio-Video Presentations, *ACM Multimedia 99*.
- Komlodi, A, & Marchionini,G. (1998). Key frame preview techniques for video browsing. In *DL'98, Proceedings fo the third ACM conference on Digital Libraries* (pp.67-75).
- Lienhart, R., Pfeiffer, S., & Effeisberg, W. (1997). Video abstracting. *Communications of the ACM*. 40(12), 55-62.
- Lindley C. A. 1997 "A Multiple-Interpretation Framework for Modeling Video Semantics", *ER-97 Workshop on Conceptual Modeling in Multimedia Information Seeking*, LA, 6-7 Nov.
- Myers, B.A., Casares,J.P., Stevens,S., Dabbish,L., Yocum,D., and Corbett,A., "A multi-view intelligent editor for digital video libraries", *JCDL'01*, June 24-28, 2001. Roanoke, Virginia, USA.
- Mu,X. & Marchionini,G.: "*Interactive Shared Educational Environment (ISEE): Design, Architecture, and User Interface*", April 2002. Technical Report, School of Information and Library Science, University of North Carolina at Chapel Hill, TR-2002-01
- Mu, X., Marchionini, G., and Pattee, A., "Interactive Shared Educational Environment: User Interface, System Architecture and Field Study" submitted to *JCDL 2003 conference*.
- Nack F., Putz, W., "Designing annotation before it's needed", 2001, *Proceedings of the ninth ACM international conference on Multimedia*, pp 261-269.
- MPEG-7, 2002, Overview of the MPEG-7 Standard, ISO/IEC JTC1/SC29/WG11 N4980, Klagenfurt, July 2002, URL at <http://mpeg.telecomitalia.com/standards/mpeg-7/mpeg-7.htm>

Srinivasan U., Gu L., Tsui K. and Simpson-Young W.G., "A data model to support content-based search in digital videos", The Australina computer journal, Vol. 29, No.4, November 1997.

Wildemuth, B., Marchionini, G., Wilkens T., Yang, M., Geisler G., Fowler B., Hughes, A., and Mu, X. "*Alternative Surrogates for Video Objects in a Digital Library: Users' Perspectives on Their Relative Usability*", 6th European Conference on Research and Advanced Technology for Digital Libraries,

September 16-18, 2002 - Pontifical Gregorian University, Rome, Italy

Wildemuth, B.M., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A., and Gruss, R., (2003) "How fast is too fast? Evaluating fast forward surrogates for digital video" submitted to JCDL 2003 conference.

Yeo, B. & Yeung, M. (1997). Retrieving and visualizing video. *Communications of the ACM*. 40(12), 43-52.