Video Learning Object Application System: Beyond the Static Reusability

Xiangming Mu and Alex Koohang
University of Wisconsin – Milwaukee
Milwaukee, WI, USA

mux@uwm.edu koohang@uwm.edu

Abstract

This paper presents an advanced an innovative Video Learning Object (VLO) application model for optimal dynamic VLO reusability. The model contains three essential elements – the metadata, the links, and a learning center. Each element of VLO application system is configured for a specific assignment. The model allows for dynamic reconfigurability of VLO metadata based on usage profile. This leads into easy adaptation of VLOs to various contexts thus attaining optimum reusability. Two multimedia application cases are presented that further demonstrate the application of the proposed model and its use for optimal dynamic reusability of VLOs. Conclusion and recommendation for future development of VLO application systems are made. `

Keywords: Video learning objects, video learning object application system, reusability, reconfigurability, dynamic reusability

Introduction

Learning objects are increasingly becoming popular in e-learning settings. A learning object is an independent piece of instruction that can be reused in various instructional contexts.

Wiley (2000) defines learning object as "any digital resource that can be reused to support learning. This definition includes anything that can be delivered across the network on demand, be it large or small. Examples of smaller reusable digital resources include digital images or photos, live data feeds (like stock tickers), live or prerecorded video or audio snippets, small bits of text, animations, and smaller web-delivered applications, like a Java calculator. Examples of larger reusable digital resources include entire web pages that combine text, images and other media or applications to deliver complete experiences, such as a complete instructional event."

The Learning Technology Standards Committee (LTSC LOM, 2002) defines learning objects as "any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning. Examples of technology supported learning include computer-based training systems, interactive learning environments, intelligent computer-aided instruction systems, distance learning systems, and collaborative learning environments. Examples of Learning Objects include multimedia content, instructional content, learning objectives, instructional software and software tools, and persons, organizations, or events referenced during technology supported learning."

Harman and Koohang (2005, p. 68) concluded that "a learning object is not

merely a chunk of information packaged to be used in instructional settings. A learning object, therefore, can include anything that has pedagogical value - digital or non-digital such as a case study, a film, a simulation, an audio, a video, an animation, a graphic image, a map, a book, or a discussion board so long as the object can be contextualized by individual learners. The learner must be able to make meaningful connections between the learning object and his/her experiences or knowledge he/she previously mastered."

Although all the above definitions include the term *video* as a learning object; Mu (2005) specifically refers to video object as a Video Learning Object with unique features such as hierarchical structure and video surrogates.

Mu (2005) advanced a VLO model of *three-tier information infrastructure* that substituted the traditional two-tier VLO structure, i.e., user interface tier and video data tier. The new model overcame the limitations of the two-tier structure, mainly the lack of standards essential for interoperability and the direct link between video metadata and the video application which limits the reusability of VLO. The three-tier information infrastructure contained *video metadata tier, video learning object tier, and video application tier.*

Today, VLOs are increasingly being used in distance education/e-learning programs. The use of VLOs is mainly for workshops, training, and education in the industry, business, government, healthcare, K-12, and post secondary educational settings. Examples of entities providing these videos for workshops, training, and education are Microsoft Multi University Research Laboratory which serves as a digital repository for universities' lectures and seminars (http://murl.microsoft.com/default.asp , Stanford Center for Professional Development (SCPD) program at Stanford University (http://scpd.stanford.edu) which provides distance education using digital videos recorded from live classes, MIT Open Courseware which is a free and open educational resource for faculty, students, and self-learners (http://ocw.mit.edu/index.html), and eduCommons, a project of the Center for Open and Sustainable Learning, is a free and open Content Management System used for Open Courseware creation and use (http://cosl.usu.edu/cosl/projects/eduCommons) among others.

Most of today's current VLO application systems are still based on the similar organization and representation strategies developed for adopting static video metadata. These systems lack suitable reusability. The static nature of the VLO's reusability makes it difficult for optimal reuse in various contexts. Even though initial efforts have been taken towards the integration of user profile and usage information (Houten, Setten, & Enschede, 2000), very little research has been done to develop innovative video organization models that overcome the limitations of current VLO application retrieval systems.

The purpose of this paper is to present an advanced model that contains essential elements, i.e., learning center, metadata, and links, leading to optimal VLO reusability. Compared to previous three-tier reusable learning object model based on metadata (Mu 2005), the new model goes beyond the static reusability. It is a dynamic model that allows for VLO's reconfigurability. Each element of VLO application system is configured to accomplish a particular task. These elements can then be reconfigured based on usage profile (e.g., who the users are, how they are using the VLOs, what the frequent links between components are, etc.) to adapt VLOs to various contexts (e.g., distance learning, video conference, collaborative project, or medical video analysis) thus achieving optimal dynamic reusability.

The paper is organized in a manner consistent with its purpose. The introductory remarks define the term learning object and the use of video as a learning object. The focus is then shifted in defining VLO and their use in distance education settings. Next, the discussion focuses upon presenting the proposed VLO Model. This presentation follows by presenting two short cases that

further illustrate the application of the proposed model. Conclusions and recommendations round out the paper.

The Video Learning Object Retrieval Model

The Video Learning Object (VLO) retrieval model is composed of three components: 1) the metadata, 2) the links, and 3) a learning center (see Figure 1).

The VLO metadata is an integration of both the concept-based (textual) and content-based (audiovisual) video metadata, including: a) video clips, b) video structure, c) video textual metadata, d) video visual metadata, and e) video audio metadata.

The VLO links refer to the internal VLO links (i.e., connections among various video components within the VLO) and external VLO links (i.e., connections between the VLO and other related VLOs.)

The VLO learning center is the mechanism enabling the VLO to dynamically update its VLO links and interface for applications that are based on the usage profile. In other words, a VLO can learn from the historical applications and changes of its content and structure accordingly.
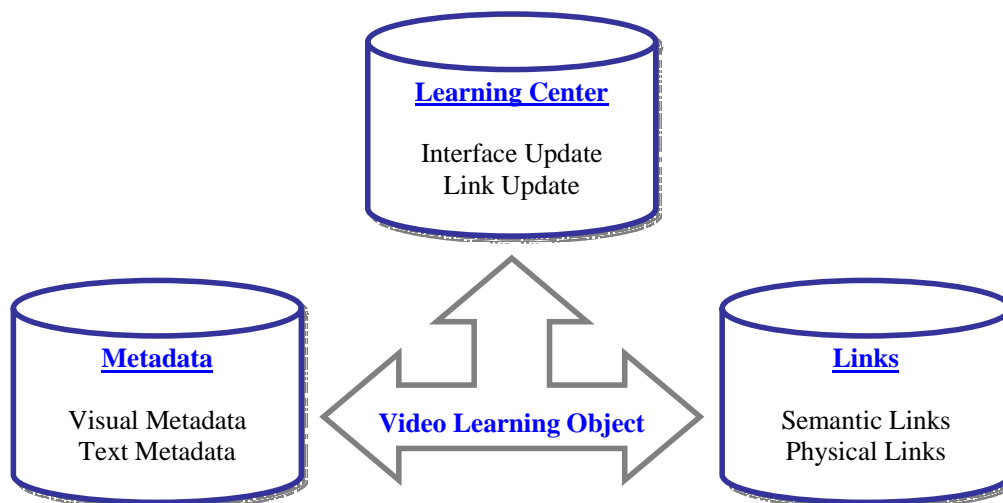
Learning Center

Interface Update
Link Update

Metadata

Visual Metadata
Text Metadata

Video Learning Object

Links

Semantic Links
Physical Links

Figure 1: VLO Retrieval System Model

The VLO Metadata

Textual and visual video metadata

The VLO Metadata contains both textual and visual Metadata. Metadata is also called "data about data" – any data that promotes the identification, description, and location of resources. This includes both physical and electronic formats. Metadata is utilized to describe the content, quality, condition, and other appropriate characteristics of the original data. For example, metadata for a book may include title, subject, description, publisher, etc.

The textual metadata for a digital video clip contains all the metadata that a book possesses (i.e., title, description, author, publisher, etc.) with additional metadata unique to the VLO, i.e., the length (how many seconds or minutes), the format (MPEG, QuickTime, etc.) and possibly other features exclusive to VLO. Textual metadata is widely used to facilitate a wide range of video

related applications, particularly in video retrieval. Video retrieval is a specific Information Retrieval (IR) process that focuses on finding video documents. One example of the video IR application is the video-based search engines. Textual video metadata are the basis for Google video engine (http://video.google.com) and Yahoo video retrieval search engine (http://video.search.yahoo.com).

Visual metadata refers to any metadata that directly relates to the visual contents of the video. Visual metadata may include: video key frames, video shots, Automatic Sound Reorganization (ASR), Machine Translation (MT), low-level camera movement features (i.e., zoom, tilt, and pan), and high-level semantic features. Naphade et al. (2005) defined a light-scale-concept ontology for the 2005 TRECVID project (http://www-nlpir.nist.gov/projects/tv2005/tv2005.html). The ontology defined seven dimensions: Program Category, Setting/Scene/Site, People, Objects, Activities, Events, and Graphics. A total of 40 individual features included in the ontology that can be utilized in the VLOs.

Visual metadata can be very helpful for content-based video information seeking and retrieval. The following illustrates visual metadata:

- *"A video that can demonstrate the structure and evolution of a category 5 hurricane" (see picture on right)*

- *"Video clips that illustrate how to effectively control the AIDS in Africa for public health training program"*

- *"A Video that demonstrates how to effectively lower cholesterol"*

- *"Video clips that teach about how to use a course management system in a distance learning setting"*

VLO metadata and IR infrastructure

Most of the current digital video retrieval systems are based on a typical two-tier infrastructure, i.e., *digital video archive tier and user interface tier*. The digital archive tier is composed of a collection of indexed digital videos and the associated metadata. The user interface tier provides support to various video related services such as video searching, ranking, and presentation. However, Mu (2005) has demonstrated that a three-tier infrastructure would improve the VLO reusability in video-based applications. For this reason, the proposed model adds a new VLO metadata tier in the traditional IR infrastructure - the VLO metadata tier bridges the traditional video archive tier and user interface tier.

Both textual and visual metadata are incorporated in the VLO metadata tier. XML markup language is utilized to code the hierarchy of the metadata. The model chooses MPEG-7 as the metadata standard. An example portion of XML code adapted from TRECVID 2005 project (http://www-nlpir.nist.gov/projects/tv2005/tv2005.html) is as follows:

<?xml version="1.0" encoding="UTF-8"?>

<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

<Description xsi:type="ContentEntityType">

<MultimediaContent xsi:type="VideoType">

```
<Video id="VLO_TEST_1">
<MediaLocator>
<MediaUri>20041116_110000_CCTV4_NEWS3_CHN.mpg</MediaUri>
</MediaLocator>
<MediaTime>
<MediaTimePoint>T00:00:00:0F30000</MediaTimePoint>
<MediaDuration>PT00H28M19S29949N30000F</MediaDuration>
</MediaTime>
<TemporalDecomposition gap="false" overlap="false">
<VideoSegment id="shot1_1">
<MediaTime>
<MediaTimePoint>T00:00:00:0F30000</MediaTimePoint>
<MediaDuration>PT00H00M03S26116N30000F</MediaDuration>
</MediaTime>
...
</Video>
</MultimediaContent>
</Description>
</Mpeg7>
```

The VLO Links

The VLO links are defined as connections between components within a VLO or connections across different VLOs. The VLO links allow different dimensions of video metadata to be integrated into a video metadata object space. In such a space, each video shot (associated with a metadata object) can be accessed from any specific dimensions: indexed words (from Automatic Speed Recognition - ASR transcript), sample image (key frame), or a specific feature (i.e., A video shot with multiple vehicles. In addition, these dimensions are linked with each other through a mechanism called "Smartlinks" proposed by Mu, Marchionini, and Patte (2003) in video collaboration project (see the example portion of XML code above). Furthermore, Christel, Hauptmann, Wactlar, and Ng (2002) demonstrated the effectiveness of the integration of related video shots retrieved along the temporal and spatial dimensions. The authors coined a word collage, which refers to the combination of text and images derived from multiple video sources, to facilitate video retrieval. The combination is based on user context, e.g.,, an originating query, coupled with automatic processing to refine the potential images.

There are two levels of links: 1) the semantic links and 2) the physical links. The semantic links are defined as connections of two or more VLOs or components with VLOs due to semantic relationships. For example, a video shot about route of the hurricane Katrina (with a map or illustration) in 2005 is linked to another shot that illustrates route of hurricane Andrew in 1992. Semantic similarities are used to decide the semantic links. Similarity between the target shot and shots in the same or different videos are calculated based on some specific criteria (the choices of the criteria is beyond the scope of this paper).

The physical links are defined as connections of the VLO components through video's time-stamps. For example, a video shot is linked to its neighbor shots through the physical links. A video script obtained from the audio channel will be linked to the associated video shot. Unlike the semantic links, the VLO's physical links are built-in. They are created automatically.

The VLO links can be created by either automatic or manual approaches. Nevertheless, they are normally created by automatic approaches. Examples of the physical links include, but are not limited to, neighboring shots, shot and its associated transcript, shot and its key frame, shot and its associated audio clip, and shot and its associated text metadata such as description, genre, length, and format. A large number of the semantic links are still created for quality consideration. The following four video shots from a video "Hurricane Force - a Coastal Perspective" (provided by Informedia Project at Carnegie Mellon University (http://www.informedia.cs.cmu.edu/team/index.html), See Figure 2) are linked with each other for their common topic of hurricane routes along the Mexico Golf coast.

In fact, automatic approaches (based on modern data mining and clustering technologies) are being applied in the creation of VLOs. For example, collages (linked videos) are used for creation of new videos Christel, Hauptmann, Wactlar, and Ng (2002).

The VLO links are closely related to the VLO metadata. A number of VLO physical links are defined as connections between different VLO metadata components. Timestamps are usually used to link video surrogates, including both visual and textual metadata, with the raw video shots. The VLO links provide "short cuts" to *direct access* of related video metadata in various video applications.
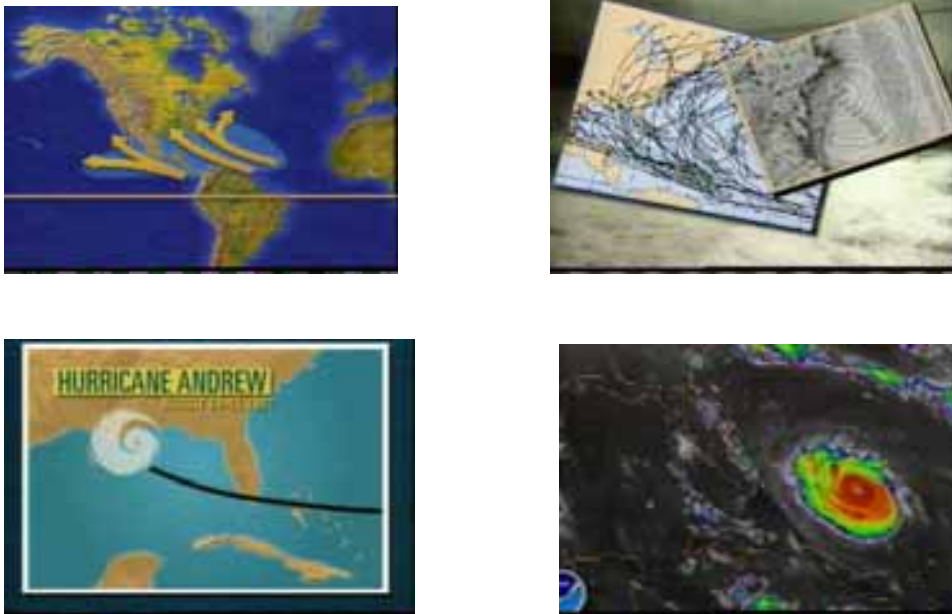
**Figure 2: Hurricane Force - a Coastal Perspective
(provided by Informedia Project at Carnegie Mellon University)**

The VLO Learning Center

The VLO learning center is a dynamic update mechanism embedded in a VLO. The VLO learning center can update both VLO metadata and VLO links based on its application profile. An application profile refers to a summary of the utilizations of a specific VLO in various applications.

A VLO learning center is an interactive *self-learning* module in the VLO retrieval system. A small database that tracks all historical applications is created inside the VLO to provide support for the VLO learning center. Each time the VLO is accessed by a specific application, the usage information will be added to the tracking database and the new information will further be utilized to update the current status of the VLO metadata and VLO links based on a pre-defined algorithm. (The predefined algorithm is beyond the scope of this paper). For example, a video search with the query "Hurricane Katrina" and the final list of user-selected video shots will lead the VLO learning center to update the metadata of those video shots to adjust the weight variables for all key words containing "Hurricane Katrina" to reflect the current status. In addition, the new application information in which a set of video shots are selected (grouped) will also trigger the VLO leaning center to adjust its semantic links to reflect the newest status. No information related to user's identification and privacy will be collected in this process.

The Cases

In this section, two short cases are presented to further exemplify the application of the proposed model and its use for optimal dynamic reusability of VLOs. Case one includes the Video Annotation and Summarization Tool (VAST) developed by Mu and Marchionini (2001). Case two uses the Interactive Shared Educational Environment (ISEE) developed by Mu and Marchionini (2002). Both cases integrated links into the traditional video learning objects to leverage the video metadata.

Case One: Implementation of VLO Model in VAST project

The Video Annotation and Summarization Tool (VAST) developed by Mu and Marchionini (2001) is a novel middleware video application environment (Figure 3). Implemented in Java which enables the interoperability across the operating systems and platforms, VAST allows users to generate, modify, and deploy video annotations on several levels: clips, segments, scenes, shots, and frames. The VLO created by the VAST provides an enriched collection of visual metadata and links. Fast-forward version of the original video clips is used as a type of video summaries. VAST extracts the frames from the raw video in an interval configured by the user and stores them in sequence which can be used directly by some tools to produce the fast-forward version of the video.

To balance the efficiency and accuracy of the key frame extracting, a unique VLO metadata authorization approach (a semi-automatic two-phase key frame authoring approach) is adopted by the VAST system. First, the automatic signal processing approach (which is not appropriate for some specific genres such as the educational "talking head" videos). Second, the manual selection approach, which is usually time-consuming, will be adapted to refine the selection of video metadata. Our strategy is to use the first phase to narrow down the selecting pool to a small number of candidates so as to dramatically improve the efficiency and effectiveness in the second manual selection stage.

Visual metadata (key frames) and semantic metadata (outlines, annotations) are both important. They are two perspectives of the same video content. Visual metadata can keep the visual gist, while the semantic metadata provides information about concepts and abstract relationships. A mapping between these two kinds of metadata is supported in VAST through the VLO links coordinated by the video timestamp. Fast forward version of the video enables user to quickly

browse the video and to grasp the visual gist effectively. A re-configuration mechanism in VAST is introduced in the VAST user interface to control the time compression ratio. The *image size* parameter can be adjusted by users to produce customized sized frames. The links were embedded in the application with the creation of video annotation. The links were automatically generated with the input of each annotation item (see two-way arrows in Figure 3).
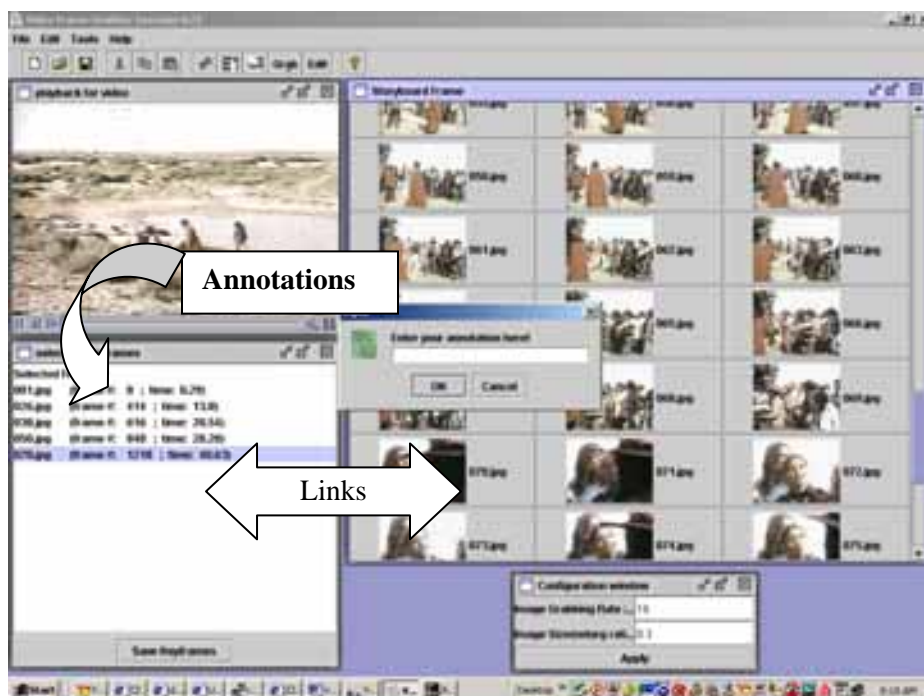

Figure 3: User interface of VAST

Case Two: Implementation of VLO Model in ISEE project

The Interactive Shared Educational Environment (ISEE) developed by Mu and Marchionini (2002) is an advanced real-time multimedia application that supports highly interactive collaboration and distance learning activities within a heterogeneous network context. The ISEE takes full advantage of fast LAN campus networks or the Internet2 Wide Area Networks to provide peer-to-peer multicast support. In addition, it can be used in the less advanced settings of home users. Media (e.g., digital video) are integrated into a desktop style interface in the ISEE. The ISEE allows users to interact with live multicasts, a shared web browser, shared video/audio with thumbnails for quick navigation, and text chat.

Collaborative work is supported via a shared time line across the multiple ISEE tools. With pre-buffering and a novel collaboration protocol, ISEE not only supports dynamic user-media interactions in real time, but also guarantees synchronization across participants.

The VLO metadata in the ISEE application (Figure 4) includes both textual and visual metadata. Different from generic VLO, hypertext documents are embedded into the VLO metadata to provide direct access through a component called "Shared Web Browser" in the ISEE interface. Other metadata displayed on the ISEE interface are the video storyboard, which is composed of a set of selected key frames.

Figure 4: User Interface of ISEE

The VLO links (see the two-way arrows in Figure 4) are widely adopted in the ISEE systems. Each of the learning and communication components in the ISEE, including Interactive Chat Room (ICR), Shared Web Browser (SWB), video player, and storyboards, are "smartly" linked through the video timestamps or hypertext URLs. In ISEE, the Interactive Chat Room (ICR) is different from the generic online text chat tools in terms of supporting interaction between users via messages. Each message sent by a user contains not only the sender's username and the message content, but also the sender's video timestamp representing the point when the message was sent. A single click on the timestamp by a particular user immediately updates that user's local video player with the same value of the timestamp. Thus any user who is interested in a particular message is able to share both the text message and the video context associated with the message. Such a mechanism is based on the VLO links and brings convenience to users who want to discuss topics that are tightly associated with the video context. For example, when discussing a video picture book, a message such as "look at the little thing sitting at the table, is it a mouse?" makes no sense to another user if he/she cannot share the same scene. For a list of current links implemented in the ISEE system refer to Mu (2005).

Integration of the Learning Center in VAST and ISEE

Both the VAST and the ISEE implemented links (primary internal links) in their VLOs, which made the VLO in both applications "smarter" than the traditional *Video plus Metadata* model. The VAST and ISEE do not possess the VLO learning center proposed in this study. Therefore, currently these systems lack the dynamic reconfigurability. Our next attempt is to integrate the learning center component in the VLOs that used for both VAST and ISEE. The learning center will leverage the application context and can be used directly by multiple multimedia applications including VAST and ISEE without introducing any extra intermediary.

Conclusion

This paper advanced a VLO application model consisting of three elements - the metadata, the links, and a learning center. The metadata contains video clips, video structure, video textual metadata, video visual metadata, and video audio metadata. The VLO links are the internal links such as connections among various video components within the VLO and the external links such as connections between the VLO and other related VLOs. The VLO learning center is the key to dynamically update the VLO links and the interface for applications that are based on the usage profile. The three essential elements work together and complement one another in a dynamic system that maximizes the reusability of VLOs in various contexts. Our model goes beyond the traditional static reusability in which the VLOs are reused in various contexts with no consideration to their adaptation to new situations. In other word, the VLOs are reused without any changes or modifications to the new contexts. The reusability is achieved, but the VLOs are not adapted to the new situation.

In our model the elements of the VLO application system are allowed for reconfigurability. Reconfigurability creates dynamic reusability. Each element of VLO retrieval system is responsible for a specific task. The elements can be reconfigured based on usage profile to adapt VLOs to various contexts (each context has its own unique situation), consequently arriving at optimal dynamic reusability.

Our model is a first stride in advancing an innovative model for VLO retrieval system that optimizes dynamic reusability. While the implementation of this model is underway by the authors of this study, it is recommended that the model be further explored. The exploration should focus upon key issues affecting the dynamic reusability of VLO retrieval systems. These issues are standards, and reconfigurability.

Standards make way for the interoperability of VLOs. There are several entities who are actively working on learning objects standards in general. Examples of these entities are IEEE LO Metadata (http://ltsc.ieee.org), Advanced Distributed Learning (ADL) Initiative - Shareable Courseware Object Reference Model (SCORM) (http://www.adlnet.org), IMS (Instructional Management System) Global Learning Consortium (http://www.imsproject.org), and the Dublin Core: Metadata for Electronic Resources (http://dublincore.org). The key questions that merit attention for standards are:

1. Which standard should be used for VLOs?

2. Will there be a common standard for VLOs?

The key questions that merit attention for reconfigurability are:

1. A VLO application system must be flexible. Can flexibility be easily embedded into reconfigurability in a VLO application system?

2. Dynamic functionality is referred to extensibility that works in concert with reconfigurability. Dynamic functionality means that functionality can be added to an implemented VLO application system without an impact on the rest of the system. What measures should one take to assure inclusion of dynamic functionality in a VLO application system?

3. While reconfigurability brings about advantages in a VLO application system, how can one address its economical feasibility?

This paper has presented an innovative model for VLO application system. The focus of this model is optimum VLO reusability. The static reusability is substituted by dynamic reusability in

which VLOs are adapted by means of dynamic reconfigurability and specifically tailored to various contexts.

References

Christel, M. Hauptmann, A. Wactlar, H. & Ng, T. (2002). Collages as dynamic summaries for news video, *MM'02*, Dec, 1-6, 2002, France.

Harman, K. & Koohang, A. (2005). Discussion board: A learning object. *Interdisciplinary Journal of Knowledge and Learning Objects*, *1*, 66-77. Retrieved from http://ijklo.org/Volume1/v1p067-077Harman.pdf

Houten, V., Setten, V., & Enschede, E. O. (2000). *Video browsing & summarization: User interaction techniques for video browsing & summarization.* Retrieved September 30, 2006 from https://doc.telin.nl/dscgi/ds.py/Get/File-12409/user_interaction.pdf

IEEE (2002). *IEEE Standard for Learning Object Metadata.* 1484.12.1-2002.

Mu, X. (2005). Decoupling the information application from the information creation: Video as learning objects in three-tier architecture. *Interdisciplinary Journal of Knowledge and Learning Objects*, *1*, 109-125. Retrieved from http://ijklo.org/Volume1/v1p109-125Mu.pdf

Mu, X. & Marchionini, G. (2001). An architecture and prototype interface for an online statistical table browser. *Proceedings of the 64th annual meeting of the American Society for Information Science and Technology*, 38, 156-170.

Mu, X. & Marchionini, G. (2002). *Interactive shared educational environment (ISEE): Design, architecture, and user interface.* Technical Report, School of Information and Library Science, University of North Carolina at Chapel Hill, TR-2002-01.

Mu, X., Marchionini,G., & Pattee,A., (2003). The Interactive Shared Educational Environment: User interface, system architecture and field study**.** *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries*, Houston, Texas, May 27-31, 291-300

Naphade, R., Kennedy, L., Kender, R.,, Chang, S., Smith, R., Over, P., & Hauptmann, A. (2005). *A light scale concept ontology for multimedia understanding for TRECVID 2005*. Retreived September 30, 2006 from http://www-nlpir.nist.gov/projects/tv2005/LSCOMlite_NKKCSOH.pdf

Wiley, D. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, in *The instructional use of learning objects.* Retrieved June 15, 2005 from http://reusability.org/read/chapters/wiley.doc

Biographies
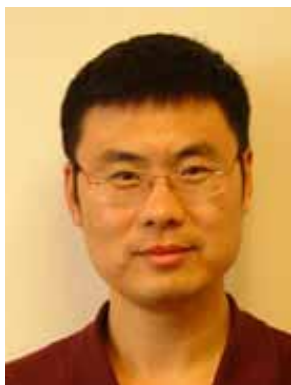
Xiangming Mu, an assistant professor from School of Information Studies, University of Wisconsin—Milwaukee, earned his PhD in Information and Library Science from University of North Carolina. He worked in multiple research projects supported by National Science Foundation (NSF) and developed a suite of new information systems: Interactive Shared Educational Environment (ISEE) to support the Video-based Collaborative Distance Learning, Browser for Enriched Statistical Tables (BEST) to help non-specialists access and use statistical data in tabular form, Video Annotation and Summarization Tool (VAST) to facilitate the creation of both semantic and visual metadata, and Table-Hunter to support context + focus view of a complex web site. His current research interests include: learning technology and learning objects, digital library, multimedia information retrieval, human-computer interaction, and data mining.

Alex Koohang is a faculty member and the director of undergraduate program in the School of Information Studies at University of Wisconsin - Milwaukee, USA. Dr. Koohang has designed, developed, and implemented various traditional, non-traditional, hybrid, and on-line courses/programs. He has been involved in the development of on-line education, having initiated and administered some of the earliest asynchronous learning networks. Dr. Koohang's current research interests are in the areas of learning objects, open education, open access, open source, e-learning, usability, and knowledge management.