SIVA Producer - A Modular Authoring System for Interactive Videos

Britta Meixner, Günther Hölbling, Florian Stegmaier, Harald Kosch
(Chair of Distributed Information Systems, University of Passau, Germany
{meixner, hoelblin, stegmai, kosch}@fim.uni-passau.de)

Franz Lehner, Martin Schmettow, Beate Siegel
(Chair of Business Administration II, University of Passau, Germany
{lehner, martin.schmettow, beate.siegel}@uni-passau.de)

Abstract: In this paper a modular authoring system for interactive videos is introduced. The system enables the author to edit the video and add annotations like images, text, hyperlinks to defined scenes or the whole video. It provides a video cutting tool to define scenes, a scene graph to realize non-linear flow of the video and a HTML editor to create text and link annotations. A project can be exported to a XML formatted file with a specific schema and Flash video format (flv) files, which can be interpreted by the project's player. The system is designed to be used intuitively.
Key Words: interactive video, interactive learning, authoring system, authoring tool, video annotations
Category: H.5.1

1 Introduction

Although most people passively watch videos on the Web, different efforts to enhance these videos with interaction features are in progress. One example for enhanced videos are YouTube Annotations. With respect to these trends and the ongoing change in entertainment, our Simple Interactive Video Authoring Suite (SIVA Suite[1]) is introduced to overcome the limits of interactive Web videos. Our system creates videos with several advantages for the viewer, like additional information about objects in a video or the ability of controlling its flow.

Definitions for interactive videos often rely on the definition by Michael K. Stenzler and Richard R. Eckert, namely that *"A video application is interactive if the user affets the flow of the video and that influence, in turn, affects the user's future choices."* [Stenzler et al. 1996]. This definition only covers the video-part itself. Recent technologies, however, offer more possibilities, like adding additional materials to well defined scenes or linking videos and homepages.

The use of interactive videos is important and can be applied to a lot of areas and scenarios. Examples of possible scenarios for realizing interactivity in videos are introduced next.

[1] http://www.siva.uni-passau.de/

Usage scenarios

Possible usage scenarios can be found among others in e-learning and knowledge management.

Many e-learning-applications make use of videos, but allow little interactivity. In many cases, it is only possible to play, stop and pause the video. Thereby it is impossible to offer an explorative and problem-oriented learning environment. Introducing interactivity can dramatically improve the usefulness of an e-learning video. If the video is split in self-defined scenes, the chronology of scenes can be customized to the knowledge of the different spectators. The user may access additional optional information, which may help him/her to understand the presented content. This information can be text, detail screens, manuals or formulas that are shown at particular parts of the video. A learning sequence that will be of interest to experts and beginners can be created. At the same time, the consumer is not put into a passive role. An authoring-system can provide options to design tests to check the learning process as a feedback to teachers and learners.

It is commonly accepted that knowledge management (KM) and knowledge transfer are critical success factors for competing enterprises. Within the discipline of knowledge management great efforts are made to develop software tools in order to support KM activities. An overview of existing tools can be found in [Lehner 2009]. A tool for creating interactive videos can be assigned to several categories used to classify KM tools (e.g. cooperation systems, content oriented systems, e-learning-systems, visualizing systems). A widely accepted and common description of KM activities is the knowledge life cycle. Core activities within the knowledge life cycle are knowledge development and knowledge generation, knowledge search and identification, knowledge diffusion, knowledge representation, knowledge evaluation and the use of knowledge. Most of these activities can be directly supported by intended functions of an authoring tool for interactive videos. For example, the annotation features can be used for cooperative knowledge development. The authoring tool will build an infrastructure and framework to manage, deliver and evolve knowledge across an enterprise with powerful features. The tool can help to make the right knowledge available to employees, partners and customers, to build knowledge capsules for specific deliverables, general product information, examples, links and any other knowledge deemed important. The collection of interactive videos can serve as a personalized knowledge repository [Lehner 2009].

Some features may be important in both scenarios. In many cases, it might be reasonable to show annotations that are related to the whole video. Examples of such annotations are names of the lecturers, the title of the whole video or links to homepages that provide more information. Switches between videos and video sequences, intersections of videos or the selection of a particular sequence

could be required, for example to show content in another perspective.

Several kinds of annotations are needed to fulfill the requirements pointed out in the previous section.

Annotations

Various types of annotations can be used in both scenarios outlined above, as well as in many other scenarios. The most basic type are simple annotations: HTML or text files, images and videos. Interactive annotations are links to websites or buttons to control the progress of the video. Marking relevant or clickable objects and slow motion brings details of video content into focus. Interactive video annotations can also be classified into three validity levels: they can be valid for the whole duration of the video (video annotations), for the period of a sequence (video sequence annotations) or refer to an object in the video (video object annotations).

Each interactive annotation can affect the main-video in different ways. When additional context gets activated, the video may stop or continue, a jump mark may be triggered or a new browser window may be opened and the video stopped. The activation can be triggered by an event or via the interaction of the user: by clicking the controls of the video, by clicking a button or a tagged object in the video or by clicking a label (link, etc.) in the information area.

The remainder of this paper is organized as follows. In section 2, an overview of several similar projects is given and the main differences compared to our proposal are discussed. Section 3 presents the SIVA Producer and section 4 highlights its architecture. A short evaluation of the used frameworks and libraries is also covered by this section. Finally, section 5 concludes the paper with a short summary and future work.

2 Related Projects

This section gives an overview of existing systems for adding annotations, features and interactivity to videos or television broadcast. Some of them provide tools to add text, images, videos and links to scenes or to mark clickable regions. It is possible to add a table of contents to navigate through the video in some systems.

Systems for interactive TV are for instance porTiVity and GMF4iTV. Both systems are inspiration for features in interactive videos but have to fulfill other requirements than these. The goal of the porTiVity-project is to develop a platform to provide Rich Media Interactive TV services with direct interactivity with moving objects [Stoll et al. 2005]. The Generic Media Framework for interactive Television (GMF4iTV) provides interactivity with moving objects and a personalized navigation within additional content [Lopez et al. 2004]. Scene-based interactivity is not possible in these systems.

Several systems for feature extraction, video sequence annotation or video navigation have been developed in the past few years. Such systems aim at finding features of a video neither than enabling an author to add material. Notable systems include M-OntoMat-Annotizer, VideoAnnEx and Vizard amongst others. M-OntoMat-Annotizer is a system for Windows platforms for extraction of visual features and association with domain ontology concepts. It has the problem of allowing no structuring of videos [Petridis et al. 2006]. The VideoAnnEx annotation system allows a simple and consistent annotation of video sequences with MPEG-7 metadata. Every shot of a video can be annotated with predefined descriptions and keywords [Smith et al. 2000]. Vizard is a video annotation system dating back to 2002, which offers content-based image indexing and retrieval and video navigation [Rehatschek et al. 2002].

Anvil is a free video annotation system for text annotations that was developed for gesture research and is suitable for other applications [Kipp 2001]. Anvil provides a graphical user interface for creating textual annotation elements. The system is platform-independent, user friendly and flexible. However images, videos and links can not be added. It is also not possible to split the video into scenes and change the flow of the video.

The academic video search engine YOVISTO[2] was developed to provide online-courses. It offers a timeline, two separate windows, text annotations and an index of the video [Sack et al. 2007]. In contrast to our system, it is neither possible to influence the flow of the video nor to add images or links.

Web-based systems can be found on the Internet. VideoClix[3] allows to add true interactivity to a video object by making it clickable. It provides an easy management and distribution of videos, but like in YOVISTO and Anvil, no influence on the flow of the video is offered. YouTube Video Annotations[4] allows the author to add background information and boxes with text to the video. Clickable areas in a video are easy to add. Adding images and HTML files is as impractical as splitting the video and changing the flow of the video.

ADIVI[5] is a system based on Flash[6] technology that offers the possibility to add aditional content (videos, links, images, .doc, .pdf) as well as areas that are tagged with actions [InnoTeams 2009]. No influence on the flow of the video can be exercised.

Virtual Campfire is a framework to create, search, and share multimedia artifacts, not exclusively videos. It offers a collection of tools called Semantic Zapping Services, Collaboration Services and Storytelling Services. The Semantic Zapping Services allow users to search and browse multimedia contents. Collaboration

[2] http://www.yovisto.com/, 26.02.2009
[3] http://www.videoclix.tv/, 26.02.2009
[4] http://www.youtube.com/t/annotations_about, 26.02.2009
[5] http://www.innoteams.de/P_ADIVI.htm, 26.02.2009
[6] http://www.adobe.com/de/products/, 23.02.2009

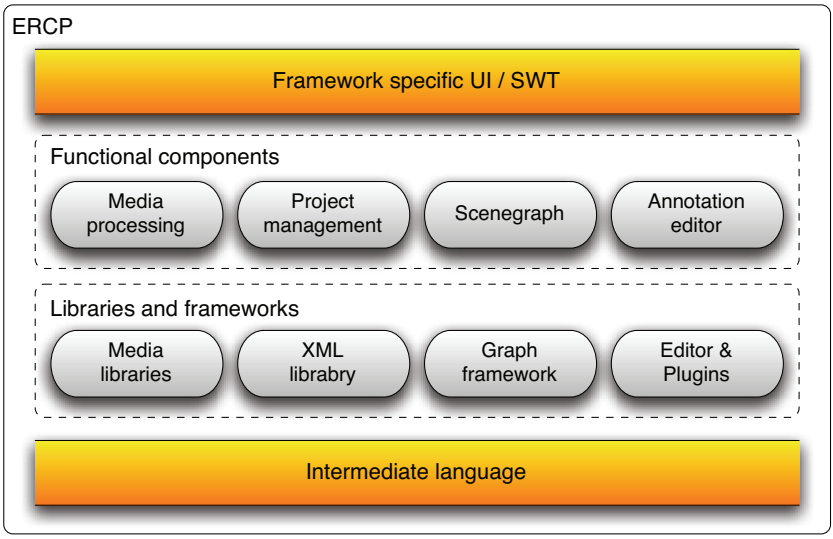

Figure 1: Architecture of the SIVA Suite Producer

Services allow users to annotate, tag and share multimedia contents within a comunity. Storytelling Services allow re-contextualization of episodic knowledge and with that a non-linearity in a story is possible [Spaniol et al. 2008].

To summarize, these authoring applications only provide a few of the necessary features described at the beginning of this section. The video flow can be rarely manipulated. Organizing the scenes e.g. in a scene graph is impossible in all these systems except Virtual Campfire. SIVA Producer offers all of the reqired features as well as a tool to define scenes, a scene-graph to manipulate the video flow and a search function.

3 SIVA Producer

SIVA Producer offers the functionality for authoring an interactive video. Several application frameworks have been evaluated as a basis for the SIVA Producer. Because of the necessity to achieve a broad support of different platforms, Java based platforms were focused. The most promising platforms were the Netbeans Platform, the Spring Application Framework and the Eclipse Rich Client Platform (ERCP). Finally, ERCP is used mainly due to its sophisticated plugin model, its comprehensive community support and its customizeable appeareance. To follow the intuition requirements, our system offers its user-interface in an operating system specific look-and-feel, because it is natively rendered by the

operating system. Users of SIVA Producer are able to customize the layout by a drag-and-drop arrangement of most UI components. SIVA Producer consists of four functional components (shown in figure 1).

– ***Media processing:*** This part of the software offers all functions needed to access, process and watch videos. Our cutting tool offers buttons and sliders for the definition of scenes, enabling an author to split a video into scenes by marking start and end position and naming the scene accordingly. To achieve better usability, a timeline with preview images of the video is used for the scene definition, as shown in Figure 2. An automated scene-detection function is integrated in the cutting tool. Furthermore, a video player with standard features is offered.

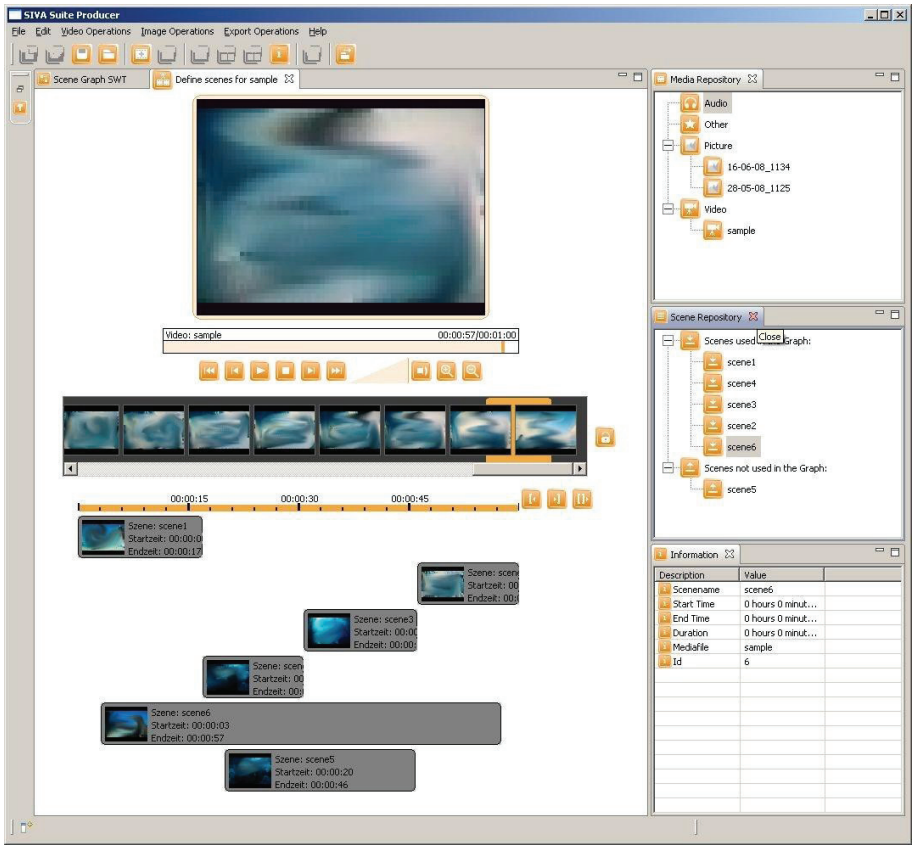

Figure 2: Media processing in SIVA Producer

- ***Project management:*** This unit offers functionality for organizing project-specific data. It helps keeping track of media files of different types, the scenes and their associated video files and additional content used in a project. E.g. available scenes and media objects are organized in repositories with a detail information area showing information about the currently selected object (see figure 2). Images, videos, HTML-files and other files can be added manually to the media repository. Scenes defined by the author are automatically added to the scene repository. Moreover, a load and save function for projects is available. After the author has finished his/her editing work, the project can be exported. This consists in transcoding all video files of the project to a selected output format (currently flv). The relations between scenes, captured by the scene arrangement in the scene graph, and their annotations are saved in an XML-document. Once the project has been exported, it is ready to be interpreted and played by the SIVA Player.
- ***Scene graph editor:*** Previously defined scenes are arranged in a scene repository from which they can be added to the scene graph by drag and drop. The scene graph allows the author to arrange the scenes in a graph that represents intersections and a non-linear flow. Scenes in the graph can be annotated with images, videos, links, buttons or HTML-files/rich text. The toolbar of the scene graph editor provides tools to delete and add nodes and edges (see Figure 3).
- ***Annotation editor:*** Offering an WYSIWYG-HTML editor allows the author to create formatted text with images and links and export the page to a HTML-file, which is added as annotation to a scene. Furthermore, predefined content saved in the media repository can also be added to a scene as an annotation.

Different libraries and frameworks are used to provide functionalities like scene editing and arrangement in the SIVA Producer. The following section describes the selected libraries and frameworks.

Libraries and Frameworks

Due to the lack of video processing support in the Java SDK and the need for such features in the SIVA Producer, media libraries for Java have been evaluated. The following evaluation criteria were used: platform independence, supported media formats, direct frame access, scene detection, expandability, documentation and support. JMF[7] combined with FOBS4JMF[8] and FFMPEG[9] was chosen as it supports the most media formats.

[7] http://java.sun.com/javase/technologies/desktop/media/jmf/, 23.02.2009
[8] http://fobs.sourceforge.net/, 23.02.2009
[9] http://ffmpeg.org/, 23.02.2009

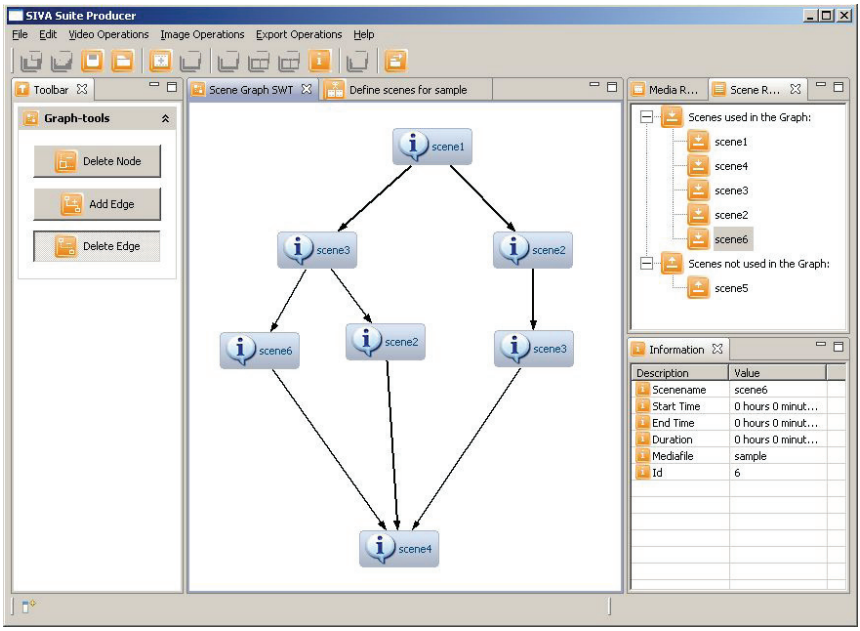

Figure 3: Scene graph editor of SIVA Producer

The library for displaying and editing graphs had to use SWT, due to the Eclipse RCP environment. We have chosen Zest[10] because it is developed by the Eclipse Foundation and uses SWT and Draw2d as GUI library. It provides layout-algorithms for the alignment of nodes. The library is available under the Eclipse Public License v 1.0 and fits perfectly into the Eclipse RCP, having been developed for it.

XML is used in many areas of our system, e.g. project management, project export, etc.. JAXB was selected as XML library for this project due to the good integration into Java, its mature status and its excellent documentation and support.

An important feature for evaluating HTML-editors is their implementation as WYSIWYG editor, to make the design of a page with images, text and links easier for amateurs. Another important feature is their availability of the code. According to volume, licensing, completeness and documentation, OnPositive's Rich Text Viewer[11] has been selected for this project. It was deployed for the Eclipse RCP and is available as a plugin. This editor is very extensive, as adaptations of the user interface can be made in the source code.

[10] http://www.eclipse.org/gef/zest/, 23.02.2009
[11] http://onpositive.com/richtext, 23.02.2009

4 SIVA Suite

The whole SIVA Suite consists of three main components - SIVA Producer (described in the previous section), SIVA Player and SIVA Server. The main purpose of SIVA Player is the playback of an interactive video provided by the SIVA Server. Therefore a player with a centered video area and additional annotation areas is offered for integration into a website. The player can be used in window mode, where all annotations are always visible, and in full screen, where annotations have to be activated by clicking a marker. An interactive video is described by an XML-document specifying the interactivity, annotations and the relationships to the main video scenes. Due to the wide-spread and broad support of Adobe Flash, video content is provided in Flash video format (flv). SIVA Player uses Adobe Flex as underlying technology. The three components interact in the following way (see Figure 4): SIVA Producer is used to add annotations and interaction to a video. The whole project is exported as a xml-file, flv-files and additional (annotation) files which are stored in a fixed folder structure. This folder structure is then stored on SIVA Server. When an interactive video is loaded, the xml file is downloaded and interpreted by SIVA Player. According to the settings in the xml file all video fragments and annotation files are retrieved at the right time to be shown in the given part of the player.

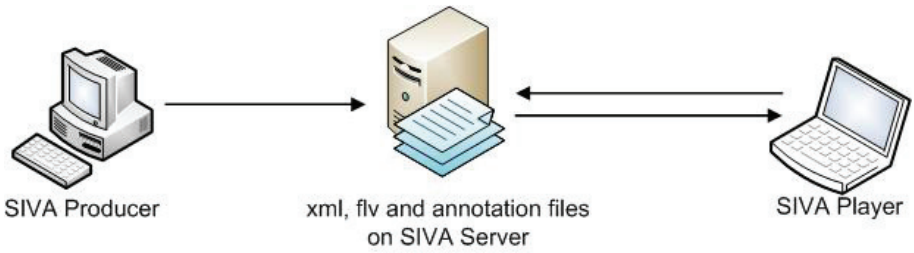

Figure 4: Scene graph editor of SIVA Producer

5 Conclusions and Future Work

The authoring system presented in this paper is designed to enable intuitive handling and to be suitable for beginners. The author can edit the video and add annotations like images, texts, hyperlinks and other videos, which will be shown in the player module. The video can be segmented into scenes by the author or via an automated scene detection. The added information is stored in a XML file for a suited display of the edited video in the player. Basic functions

of the authoring system are creating, loading and saving a project, importing video and media, adding, deleting and playing media. Another basic function is splitting the videos into scenes, which can be named, arranged and annotated in a scene graph. Finally, the whole project can be exported.

A prototype of the system with basic functionality is available. Next work concerns a scene detection. Future tasks are object marking and object tracing and a well developed scene graph with zoom functions and other features that facilitate working with large graphs. An important task in future work is the definition of a reliable and flexible file format to provide seamless cooperation with the SIVA Player.

Acknowledgments

This work was partually supported by European Social Fonds and the Bayrisches Staatsministerium für Wissenschaft, Forschung und Kunst (Bavarian State Ministry of Sciences, Research and the Arts) under project name “iViPro”.

References

[InnoTeams 2009] InnoTeams GmbH: ADIVI System Handbook `http://adivi.meticube.com/documentation/pdf/adivi_system_handbook.pdf` (20.02.2009)

[Kipp 2001] Kipp, M.: “ANVIL A Generic Annotation Tool for Multimodal Dialogue”; In: Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech), Aalborg, September 2001, 1367-1370

[Lehner 2009] Lehner, F.: Wissensmanagement - Grundlagen, Methoden und technische Unterstützung. Hanser Verlag, München, 3. Auflage 2009

[Lopez et al. 2004] Alejandro López, Miquel Noè and Gabriel Fernàndez: “Linkage of additional contents to moving objects and video shots in a generic media framework for interactive television”; presented at SPIE Conference, Optics East / ITCom 2004, Philadelphia, October 2004

[Petridis et al. 2006] Petridis, K., Anastasopoulos, D., Saathoff, C., Timmermann, N., Kompatsiaris, Y., and Staab, S.: “M-OntoMat-Annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Features”; Engineered Applications of Semantic Web Session (SWEA) at the 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES 2006), Bournemouth, U.K., October 2006

[Rehatschek et al. 2002] Rehatschek, H., Kienast, G.: “VIZARD - an innovative Tool for Video Navigation, Retrieval, Annotation and Editing”, proceedings of the 23rd workshop of Plattform fr verteilte Anwendungen (PVA 2001) ’Multimedia and Middleware’, Vienna, May 2001

[Sack et al. 2007] Sack, H. and Waitelonis, J.: “OSOTIS-kollaborative, inhaltsbasierte Video-Suche”; in Proc. of 5. e-Learning Fachtagung Informatik, DeLFI 2007, Siegen, September 2007

[Smith et al. 2000] Smith, J. R. and Lugeon, B.: “A Visual Annotation Tool for Multimedia Content Description”; Proc. SPIE Photonics East, Internet Multimedia Management Systems, November 2000

[Spaniol et al. 2008] M. Spaniol and R. Klamma and Y. Cao: Learning As a Service: A Web-Based Learning Framework for Communities of Professionals on the Web 2.0; Advances in Web-Based Learning, Proceedings of ICWL 2007, Edinburgh, Scotland, August 2008, 160-173

[Stenzler et al. 1996] Stenzler, M., Eckert, R. (eds.): "Interactive Video"; SIGCHI Bulletin Vol. 28 No. 2, April 1996, 76-81.

[Stoll et al. 2005] Gerhard Stoll et al.: "porTiVity: new Rich Media iTV Services for handheld TV"; 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, London, December 2005