Hypervideo and Annotations on the Web

Madjid Sadallah
DTISI - CERIST
Alger, Algérie
Email: msadallah@mail.cerist.dz

Olivier Aubert
LIRIS - Université Lyon 1
UMR 5205 CNRS
Email: olivier.aubert@liris.cnrs.fr

Yannick Prié
LIRIS - Université Lyon 1
UMR 5205 CNRS
Email: yannick.prie@liris.cnrs.fr

***Abstract*—Effective video-based Web information system deployment is still challenging, while the recent widespread of multimedia further raises the demand for new online audiovisual document edition and presentation alternatives. Hypervideo, a specialization of hypermedia focusing on video, can be used on the Web to provide a basis for video-centric documents and to allow more elaborated practices of online video. In this paper, we propose an annotation-driven model to conceptualize hypervideos, promoting a clear separation between video content/metadata and their various potential presentations. Using the proposed model, features of hypervideo are grafted to wider video-based Web documents in a Web standards-compliant manner. The annotation-driven hypervideo model and its implementation offer a general framework to experiment with new interaction modalities for video-based knowledge communication on the Web.**

I. Introduction

Interactive video is an augmented and uncommon form of digital video that accepts and responds to the input of a viewer beyond the conventional interactive features [1]. While this concept addresses a wide and growing range of online digital consumers, effective use of interactive video still rises many unresolved concerns. Indeed, the continuous and time-dependent nature of video results in information distributed over space and time, making it more difficult to apprehend.

Hypervideo is a kind of hypermedia document that brings powerful capabilities to video-centric documents with more elaborated practices and advanced interactivity. In the present work, an annotation-driven hypervideo model is proposed and a practical framework for hypervideo creation and rendering is developed, building upon video features and Web facilities to illustrate online interactive video usage.

The article is organized as follows. Section 2 introduces hypervideo field. In Section 3, the motivations for annotation-driven hypervideo are discussed. Section 4 presents the Component-based Hypervideo Model (CHM) and its annotation-based principles. In Section 5, an implementation is presented along with an illustrative example. Section 6 discusses the proposed model and implementation.

II. Hypervideo

A. Definition

Commonly, videos are embedded within hypermedia documents as atomic, sequential, static and not easily navigable clips mainly used as explanation support to give a better idea of a concept [2]; hence, they are not in the core of the hypermedia system. Hypervideo addresses this issue by a specialization of hypermedia centered on interactive video. We define hypervideo as being an interactive video-centric hypermedia document built upon an audiovisual content - a set of video objects -. Several kinds of related data are presented within the document in a time synchronized way to augment the audiovisual part around which the presentation is organized in space and time.

The articulation of video content with navigation facilities introduces new ways for developing interfaces, rendering the content and interacting with the document. By supplying spatio-temporal behaviors to links, hypervideos allow addressing and deep linking video details in order to fully integrate video to the document hyperspace. The time-based nature of hypervideos, combined with the complexity originated from conventional hypermedia involves different aesthetic and rhetoric design issues [3], [4] with regard to the possible user attitudes towards the nature of these documents.

A high degree of flexibility is granted by hypervideos to compose video-based documents and to define means and tools that foster the carried knowledge perception by promoting an active reading experience and highlighting the audience dynamic engagement and influence on the document storyline [5]. This is achieved through the use of the structure and the internal organization exposed by hypervideos to define advanced features like audiovisual montage and inter-connection, and synchronous/simultaneous display of different videos. Hence, many visualization and navigation paradigms supply the document presentation: attaching a table of content to the video, navigation across video scenes and objects, synchronous display of textual - subtitles, comment, etc.- and graphical overlays, and presenting the content following different points of view. Moreover, hypervideos allow the presentation to shift from one visualization mode to another through rich and user-friendly interfaces. For instance, while a document presents a video supplemented with a table of content, a user may activate a link to see more details about a feature and thus, the presentation may shift towards presenting only a video player accompanied by textual and graphical overlays.

B. Use Cases

The large applicability spectrum of hypervideos covers the numerous areas that already use video as a means for knowledge sharing. While very early proposals for hypervideo targeted new kinds of cinematic experience [3] by offering filmmakers means to define different narrative sequences and viewers the choice of the path to follow, novel usages are experienced nowadays.

Hypervideo movies analysis is at the core of many projects that aim to develop means to navigate in movies through visual representation of content and structure. It is for instance possible to reveal the movie structure and to enrich its content but also to compare sequences or to link and render many movies in a unified scenario. In training and scholar systems, hypervideo supports the creation of rich and realistic learning environments [6] by offering dynamic visualizations and facilitating reflective learning and cognitive flexibility [7]. Greater control and autonomy are granted to the learner to explore video data complemented with external materials and presentation/navigation aids like maps and history.

Some hypervideo providers focus on the interest in multimedia-based marketing to allow the design of online video clips that link to advertising or e-commerce sites, or to provide more information about particular products. Hypervideo can be found in many other fields: entertainment programs such as video games where users can participate in the presented movie; interactive documentaries where the author may recombine video segments and expose an iconic representation of some video details enriched by an extra content; and interactive television where additional and interactive content is synchronized with a live transmitted program [5].

C. Related Works

HyperCafe [3] is one of the earliest video-centric hypermedia research projects that discusses general hypervideo concepts. Recent projects like Hyper-Hitchcock [8], Advene [9], HyLive [5], HVet [6] and SIVA Suite [10] allowed the emergence of new hypervideo practices making it possible to edit and present hypervideos with advanced features.

With online video sharing, the growth of mobile devices and video tools, the new multimedia practices are bringing hypervideo ideas to wider usages. Indeed, interactive video is increasingly becoming a dominant feature of online media platforms. Thanks to popular platforms like YouTube annotations, audiovisual documents get more interactive by integrating graphical annotations in a video and generating rich presentations. Online hypervideo editors like *Asterpix*[1] and *VideoClix*[2] combine video, social media, interactivity and commerce in one experience. *Popcorn.js*[3] is an experimental tool for connecting video to the rest of the web, linking it into the hypertext world, pulling data from Google Maps, Twitter and Wikipedia and synchronizing it with video documents.

Most of hypervideo authoring and reading tools use specific approaches for abstracting their documents with implementation based representations, resulting in informal models. For instance, Hyper-Hitchcock provides an environment based on a very restrictive form of hypervideo. Other model-based systems consider hypervideo from very general hypermedia perspectives empowered with some specific features like video hyperlinks. Technologies like SMIL [11], NCL [4] and Flash [5] are some of the languages and tools used for implementing hypervideo engines.

The design of hypervideos by adding structure and content to the video emphasizes the importance of metadata and annotations in the process of creating such documents. Annotation-based approach for hypervideo composition offers many interesting features for document edition, presentation, maintenance and exchange. While many systems support adding content and structure to video material, very few use annotations as a key concept for document generation.

III. Video Annotations and Hypervideos

A. Video Annotation

The nature of video data is unsuitable for traditional forms of data access, indexing, search, and retrieval [12]. Moreover, audiovisual-based documents raise many concerns regarding data rendering, linking and navigation from and to precise parts, enriching and explaining the content, arranging and revealing the story structure, etc. This motivated efforts towards the use of annotations in the disclosure, explanation and augmentation of the knowledge carried by the video. Annotation-based modeling of audiovisual documents adds a new content layer, depicted by annotations, on top of video documents [13] to define the needed infrastructure for augmenting the medium with meaningful data. This data can be further used to manage and manipulate videos for analyzing, indexing, searching and generating enriched views of the document.

We consider in this article an annotation as a piece of data associated to a video fragment [9], which is a logical video segment defined by the starting and ending timecodes. Many dedicated environments, such as Advene [14], Anvil [15] and IBM's VideoAnnEx[6] offer audiovisual document annotation. They may use various formats for storage, but the MPEG-7 [16] standard offers a comprehensive and interoperable set of audiovisual description tools to represent such metadata.

B. Towards Annotation-based Hypervideos

The use of video annotation to generate video-based hypermedia document visualizations is often referred to as hypervideo, since it breaks the linearity of video and creates a non-linear information space [17].

While many works targeted the use of annotations for audiovisual document description and enrichment, the annotation structure is often tightly linked with the video when it is not completely embedded in the stream or in the player. Some systems like Hyper-Hitchcock [8] define annotation as key concept for hypervideo documents design. However, their approaches for annotation definition does not make use of any standard format or explicit data model [18] and the implied uses and representations are mainly technically driven. Other

[1]http://www.asterpix.com/
[2]http://www.videoclix.tv/
[3]http://popcornjs.org/
[4]http://www.ncl.org.br/
[5]http://www.adobe.com/products/flashplayer/
[6]http://www.research.ibm.com/VideoAnnEx/

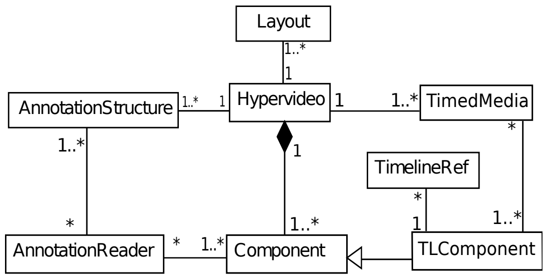

Fig. 1. CHM General Overview

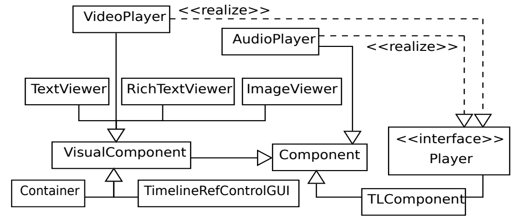

Fig. 2. CHM Plain Components

systems like Youtube use annotations mainly as captions, branching anchors and graphical overlays over the video, while a clearer separation between content and representation could foster innovative uses. Therefore, as these systems only provide a few of the necessary support for defining annotation-based hypervideo data models, the current use of annotations prevents the emergence of fully annotation-based techniques for hypervideo design. Our Advene project [9] is an effort to develop such a model and to propose solutions for annotating videos and generating hypervideos.

IV. Annotation-driven and Component-based Hypervideo Model

A. Rationale

We intend to propose an annotation-driven data model that allows the generation of multiple hypervideo presentations, and thus visualizations of the same content, from existing annotation structures. The component-based approach proposes a logical view of the document that tries to be accessible and expressive, presented as a hierarchy of components. Such an approach eases the conceptualization and coding tasks of hypervideos and allow the model to be extensible since new components can be created from existing lower-level ones. The proof of principle of the model concepts and assumptions is illustrated by a Web implementation.

B. CHM General Overview

CHM, which stands for *Component-based Hypervideo Model*, is a an annotation-driven and component-based data model to conceptualize hypervideos and describe their main features. A general overview of the model is provided in figure 1. CHM uses the principle of nesting high level components for document hierarchical composition.

A CHM hypervideo is composed of a set of low and high level components, building blocks that represent formal information and composition units. These components are related to video streams referenced by *TimedMedia* components. A TimedMedia component has an intrinsic duration which, when played through a player component, conveys a timing capability to the document, expressed by a virtual reference - an abstract clock - called *TimeLine Reference* (TLR) that synchronizes the related rendering components. Many players (therefore, many TLRs) may be present within the same document, defining different hypervideo sub-documents possibly spatially and/or temporally related. This allows synchronization between numerous hypervideo sub-documents and to use one to enrich another. Hence, temporal constraints may exist between the different TLRs.

The dynamic behavior of a CHM hypervideo is handled by an *event-based* mechanism. An event is triggered by a particular environment change and may be associated a set of *actions*. An action can be an atomic instruction that acts on the document or a set of operations which may trigger other events and cause further actions.

C. CHM Rendering Components

A *Hypervideo* consists in a set of rendering *components*, associated with a list of composition, placement, synchronization and behavioral attributes supplied by the author or retrieved from the annotation structure. A hypervideo references at least one main audiovisual document accessed through the `TimedMedia` element that addresses a temporalized stream, audio or video, to define a timeline reference (TLR).

A generic `Component` within a hypervideo may not relate to any TLR, and if so is said to be time-independent. Components bound to a TLR are specialized `TLComponent` elements with synchronization constraints. A global timeline reference allows synchronization between different TLRs.

Content with visual manifestation is displayed by `VisualComponent` elements. Synchronized *display components* offer interactive interfaces for rendering temporalized data, provided as annotations. Figure 2 presents some of these plain data components. They cover static content viewer (`TextViewer`, `RichTextViewer` - container for heterogeneous content like HTML pages, RSS feeds and broad XML-based content - and `ImageViewer`), continuous media players (`VideoPlayer` and `AudioPlayer` that present a generic `Player` interface for rendering and interacting with the content) and containers (`Container` elements) for spatial clustering of components to unify their processing. The `TimelineRefControlGUI` defines a graphical user interface for controlling and interacting with TLRs. These plain components are combined in order to propose the set of high level components shown in figure 3.

The `History` component is a graphical representation of the user interaction history, allowing to save and to restore the document presentation context by tracking the user navigation.

A document map (`Map` component) is a general sketch - made of *screenshots* or *transcriptions* - of the presented content that offers a branching opportunity choice to navigate

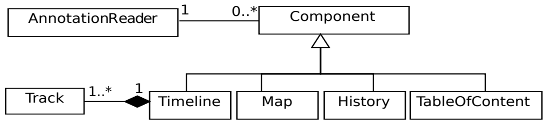

Fig. 3. CHM High Level Components

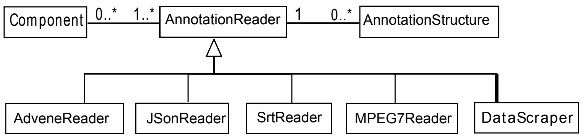

Fig. 4. CHM Data Access Components

to a particular perspective. A hypervideo table of content (`TableOfContent` (ToC) component) provides navigation capabilities and reveals the structure of the video according to a selected feature or annotation type. Many maps and tables of contents may be instantiated within the document, following different presentation styles.

Graphical, chronological and interactive representation of the hypervideo time is provided by the `Timeline` components. They place media elements, meaningful annotations and links along a timed axis, on different tracks (`Track` elements).

D. Annotation-based Model

The CHM annotation concept is defined as any information - data or resource - associated with a logical spatio-temporal video *fragment*, that defines its scope - begin and end instants - in video time and its coordinates relatively to the video frames. For instance, a spatio-temporal anchor is defined by an annotation that addresses a fragment covering its presentation interval. A fragment is thus defined as a logical representation of a video segment - not physically extracted - addressed by its timecodes. Attributes of an annotation include its *type*, *media reference*, *begin/end timecodes* and *content*. Depending on user requirements, the proposed annotation attributes can easily be extended and adapted to specific needs. While annotations add structure and content for audiovisual documents, CHM relies on them in many ways to generate hypervideos and to supply them with interaction and visualization artifacts like anchors/hostpots and links, overlays, subtitles, comments and tables of content.

E. Data Access Components

Data access components presented in figure 4 are middleware components with functional interfaces that offer unified access to the data structure (annotations and resources). Multiple readers can provide annotation data, either user-supplied or possibly automatically extracted from media elements (textual transcription, screenshots, videoclip references, etc.). The `AnnotationReader` element describes the generic data access interface. Among the dedicated data readers: (1) the `JSonReader` element allows reading from a JSon[7] file, (2) the `SrtReader` element allows using *srt* files for timed text, (3) the `MPEG7Reader` element gets data from MPEG-7 files, (4) the `DataScraper` components generate annotations by scraping DOM subtrees (from the current document or from external ones) and (5) the `AdveneReader` is the data access component for annotation performed with the Advene annotation tool, which is used for reading data from an explicit annotation structure.

[7]http://www.json.org/

F. Spatial, Temporal and Link Models

The visual components of a CHM hypervideo are placed within *spatial regions* that form the document *layout*. The placements and dimensions can be expressed explicitly or implicitly, absolutely or relatively. Presentation specification attributes are associated to the components, regions and layout and can be used by the rendering engine.

The document temporal specification is achieved through a timeline-based model and relies on the *Timeline Reference* (TLR) concept. Time-based components are activated/deactivated when reaching specific timecodes provided or computed by reference to the TLR. The non time-dependent objects are associated with the global document clock, the top level reference of the entire user-declared TLRs. The access and control of a TLR is performed thanks to the “position”, “state” and “duration” attributes. Any update of the TLR position or state affects all the related component playback.

CHM hypervideo links are defined in space and time and are unidirectional, thus they are attributes of the source anchors. A classic hypertext anchor can be defined on a specific region of a textual or graphical component. When placed on a region of a continuous media, with spatial and temporal constraints and eventually within a moving region - whose location changes over time -, the anchor is called *hotspot*.

Links may be internal or external. Activating an internal link causes a temporal shift of the presentation by an update of the TLR position. A return behavior can be specified to express whether the presentation will pause, stop, return to the source anchor point or continue from the target one. An external link leads to a foreign anchor expressed by an URL and causes the presentation to pause, continue or stop playing.

V. Web-based Model Implementation

Based on CHM, a syntax for authoring hypervideos is proposed as an extension above the HTML language. CHM namespaced attributes are associated to standard HTML elements, and extend them with specific behaviors. Such behaviors are handled by a set of generic and extensible JavaScript libraries that perform content transformation to dynamically generate standard HTML code, readable and understandable by modern browsers. Complex hypervideos can therefore be authored as standard Web documents, styled with CSS, extended by SVG and controlled by scripts. Common Web content is written in standard HTML while the hypervideo components are expressed through the CHM attribute-based syntax. For rendering the audiovisual content, we use the latest HTML standard proposal [19] `<video>` element and its corresponding DOM API specification.

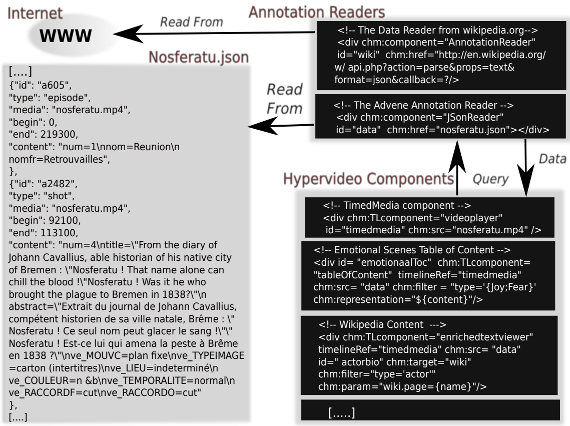

Fig. 5. Data Access by Annotation Query

Annotations can be made by hand or using dedicated tools. The defined structure has to ensure the definition of the model attributes, especially timecodes for fragment definition. A fragment is addressed with a syntax that contains the target stream along with begin or begin/end instants. We are using W3C MediaFragment [8] proposal to address temporal and spatial media fragments. For instance *http://www.exemple.com/video.mp4#t=.120,1.230* addresses the fragment of the video identified by its begin and end timecodes. When the end instant is not defined, this supposes it to be equal to the medium duration.

The CHM spatial model is supported by the HTML layout model via cascading stylesheets (CSS) and container elements like `div`. Document time and element synchronization is handled by a JavaScript implementation of SMIL Timesheets (External Timings) [20], one of the interesting addition to SMIL to make SMIL 3.0 [11] timing and synchronization available in other XML-based markup languages. Most of Web standards-based video projects like *popcorn.js* exclusively use the time inferred from the video stream to schedule the presentation. In our prototype, the timing mechanism allows hypervideo to use multiple time references, thanks to the timeline-like temporal model implemented by SMIL Timesheets. A first version of the proposed language and tools has been developed[9], and further developments are underway.

Example: To provide some features revealed by the Nosferatu movie [10] analysis performed with the Advene tool [11], a hypervideo document is designed - its rendering is shown in figure 6 - , based on the annotation package produced by Advene. Once the generated package is available into JSon format, the CHM implementation syntax is used to design rich views of the movie. Many CHM display components are instantiated: data readers, enrichment content viewers, video players, hotspots, timelines, maps and tables of contents. As shown in figure 5, the hypervideo makes use of data provided by the JSon package and possibly from the *Wikipedia.org* encyclopedia, in a synchronized way. To retrieve the desired data, two readers are defined, parameterized by a query string: a JSon reader - to query movie annotations - along with a generic annotation reader to fetch Wikipedia documents. For instance, the display component that synchronizes biographic information about actors when they appear on the movie uses both readers. First, the identity and the timecodes of each actor appearance are retrieved through the JSon reader and the received data is used to pull information from *Wikipedia*. The final data is displayed by the component according to its specifications in style and time.

[8]http://www.w3.org/2008/WebVideo/Fragments/
[9]See http://www.advene.org/chm/ for code and examples.
[10]http://www.archive.org/details/nosferatu
[11]http://www.advene.org/

VI. Discussion

With the CHM high level concepts, hypervideo definition and implementation are quite straightforward. The annotation-driven approach defines various access granularities to video, by allowing fragment definition with no length restriction. Fragments can be arbitrarily overlapped or nested, associating various data and multiple views to the same audiovisual information unit. By separating data from content presentation, maintaining the document structure is eased, independently from the audiovisual stream. This enforces security management and enhances collaboration options by requiring only the annotation structure update or exchange. The annotation structure definition in CHM is system-independent though a minimalistic schema is provided. The model realization may be achieved by any annotation standard such as MPEG-7 or a user-defined structure.

While a lower abstraction representation of the model can be deduced using common hypermedia models, CHM strength resides on its high level abstraction and its usage expressiveness. Being hypervideo specific, we claim that it is general enough to describe the existing domain features and to conceptualize the majority of existing hypervideo documents.

The model has been implemented to practically experience and evaluate our claims. To better theorize interactive video-centric documents on the Web, considered as annotation-based hypervideos, a first implementation is developed by grafting hypervideo specific concepts and attributes in a declarative manner to the HTML language. This makes authoring such documents more convenient compared to the existing systems, thanks to the Web standards-compliant syntax; writing a hypervideo document is as easy as editing any conventional Web page. SMIL presents a robust and comprehensive model and tools for multimedia/hypermedia documents composition. However, authoring hypervideo documents with the very general concepts introduced by SMIL and its huge language specification - not widely implemented by common browsers - is an arduous task.

VII. Conclusion

In this paper, we have presented an annotation-driven approach for defining a component-based hypervideo model. To theorize hypervideo, the proposed model promotes a clear separation of data from the rendered content, through the use

Fig. 6. Nosferatu Hypervideo Example

of annotations and high-level rendering components. We have used this model to propose an implementation that exemplifies enriched video support on the Web, by grafting hypervideo specific concepts and attributes in a declarative manner and a Web standards-compliant syntax to the HTML language. Future works include the implementation of the whole model specification and the experimentation of more use cases and examples. The authoring process may be eased by a graphical user interface (GUI) environment, this is one of the future development stage. We want also to rely on cognitive theories to propose more meaningful rendering components and libraries and other interaction possibilities like supporting end-users annotations and hypervideo generation.

Acknowledgments

This work has been partially funded by the French FUI (Fonds Unique Interministériel) - CineCast project and the ANR (Agence Nationale de la Recherche) - Ithaca project. It is supported by the Algerian Research Center on Scientific and Technical Information (CERIST).

References

[1] R. I. Hammoud, "Introduction to interactive video," in *Interactive Video*. Springer-Verlag, 2006, pp. 3–24.

[2] T. Navarrete and J. Blat, "VideoGIS: Segmenting and indexing video based on geographic information," in *5th AGILE Conference on Geographic Information Science*, 2002, pp. 1–9.

[3] N. N. Sawhney, D. Balcom, and I. E. Smith, "HyperCafe: Narrative and Aesthetic Properties of Hypervideo," in *UK Conference on Hypertext*, Bethesda, Maryland, United States, 1996, pp. 1–10.

[4] T. Chambel, C. Zahn, and M. Finke, "Hypervideo and cognition: Designing video-based hypermedia for individual learning and collaborative knowledge building," in *Cognitively Informed Systems: Utilizing Practical Approaches to Enrich Information Presentation and Transfer*, E. A. (Ed.), Ed., 2006, pp. 26–49.

[5] P. Hoffmann, T. Kochems, and M. Herczeg, "Hylive: Hypervideo-authoring for live television," in *Proceedings of the 6th European conference on Changing Television Environments*, ser. EUROITV '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 51–60.

[6] C. A. Tiellet, A. G. Pereira, E. B. Reategui, J. V. Lima, and T. Chambel, "Design and evaluation of a hypervideo environment to support veterinary surgery learning," in *21st ACM conference on Hypertext and hypermedia*, 2010, pp. 213–222.

[7] C. Zahn, B. Barquero, and S. Schwan, "Learning with hyperlinked videos - design criteria and efficient strategies for using audiovisual hypermedia," *Learning and Instruction*, vol. 14, pp. 275–291, 2004.

[8] A. Girgensohn, L. Wilcox, F. Shipman, and S. Bly, "Designing affordances for the navigation of detail-on-demand hypervideo," in *Proceedings of ACM Advanced Visual Interfaces*, May 2004, pp. 290–297.

[9] O. Aubert and Y. Prié, "Advene: active reading through hypervideo," in *Proceedings of the 16th ACM conference on Hypertext and hypermedia*. New York, NY, USA: ACM, 2005, pp. 235–244.

[10] B. Meixner, B. Siegel, G. Hölbling, F. Lehner, and H. Kosch, "SIVA suite: authoring system and player for interactive non-linear videos," in *Proceedings of the international conference on Multimedia*, ser. MM '10, 2010, pp. 1563–1566.

[11] D. Bulterman, "Synchronized Multimedia Integration Language (SMIL 3.0)," W3C, W3C Recommendation, 2008, http://www.w3.org/TR/SMIL3/.

[12] J. Li, Y. Ding, Y. Shi, and J. Zhang, "Building a large annotation ontology for movie video retrieval," *JDCTA*, vol. 4, no. 5, pp. 74–81, 2010.

[13] R. Fagá, V. G. Motti, R. G. Cattelan, C. A. C. Teixeira, and M. da Graça Campos Pimentel, "A social approach to authoring media annotations," in *10th ACM symposium on Document engineering*, ser. DocEng '10, 2010, pp. 17–26.

[14] O. Aubert, P.-A. Champin, Y. Prié, and B. Richard, "Active reading and hypervideo production," *Multimedia Systems Journal, Special Issue on Canonical Processes of Media Production*, p. 6 pp., 2008.

[15] M. Kipp, "Anvil - a generic annotation tool for multimodal dialogue," 2001.

[16] J. M. Martinez, R. Koenen, and F. Pereira, "MPEG-7: The generic multimedia content description standard, part 1," *IEEE Multimedia*, vol. 9, no. 2, pp. 78–87, 2002.

[17] G. Miller, S. Fels, A. A. Hajri, M. Ilich, Z. Foley-Fisher, M. Fernandez, and D. Jang, "Mediadiver: Viewing and annotating multi-view video," in *Computer Human Interaction; Interactivity Session*, May 2011.

[18] T. Alisi, M. Bertini, G. D'Amico, A. Del Bimbo, A. Ferracani, and G. Serra, "Arneb: a rich internet application for ground truth annotation of videos," in *Proc. of ACM International Conference on Multimedia (MM)*. ACM Press, October 2009.

[19] I. Hickson, "HTML5," http://www.w3.org/TR/html5/, January 11, 2011.

[20] F. Cazenave, V. Quint, and C. Roisin, "Timesheets.js: When SMIL meets HTML5 and CSS3," in *to appear in Proceedings of DocEng2011*, 2011.