Advene : active reading through hypervideo

Olivier Aubert, Yannick Prié

Advene: active reading through hypervideo

Olivier Aubert
LIRIS UMR 5205 CNRS
Lyon 1 University
F-69622 Villeurbanne, France
olivier.aubert@liris.cnrs.fr

Yannick Prié
LIRIS UMR 5205 CNRS
Lyon 1 University
F-69622 Villeurbanne, France
yannick.prie@liris.cnrs.fr

ABSTRACT

Active reading and hypermedia usage are an integral part of scholar daily practices, but the full exploitation of their potentialities still lies far ahead. In the search for new methods and tools, we focus in this article on the use of audiovisual material in a scholar context. One of the results of active reading applied to audiovisual material can be hypervideos, that we define as views on audiovisual documents associated with an annotation structure. The notion of hypervideo is useful to analyse existing video-based hypermedia systems as well as building new systems. The Advene project proposes an implementation of hypervideos through a framework that allows experimentations of new visualisation and interaction modalities for enriched videos.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Video; H.5.4 [**Hypertext/Hypermedia**]: Architectures, Navigation

Keywords

Annotation, Advene, Hypervideo, Document template, Video Information Visualisation, Sharing, Time and synchronisation

1. INTRODUCTION

The vision of the Memex proposed by Vannevar Bush has slowly matured over the years, and to a certain extent become a widespread reality with the Web. A part of the original ideas has been implemented, when tools and technology allowed it. It has also taken a while for people (be they scholars or personal users) to begin to apprehend the nature of hypertext, and devise new interaction practices with this new medium. People are currently still learning to work with hypermedia documents, both in terms of production (authoring methods and tools) and in terms of reading (practices and tools [15]).

HT'05, September 6–9, 2005, Salzburg, Austria.

Scholars were among the first concerned by the introduction of hypertext and the Web, which finds its roots in the intent of Berners-Lee to allow the CERN researchers to easily share their results. Scholar usage of hypertext notably presents the following characteristics: linking, interaction and collaboration.

Linking (from Bush's *trail of association* to the Web) is in the nature of any scholar work. It has however been argued that hypertext brings a whole new dimension and does not sum up to a simple linking of existing material. Kolb [17] for instance offers an interesting reflection about the potentialities offered by hypertext for philosophical works. The *interaction* of the user with the material is increased, in terms of manipulation of the medium (document annotation, automatic extraction of information based on queries, etc.) and in terms of navigation (bookmarks, navigation history). *Collaboration*, by the sharing of information, is at the roots of scholar works, and also a key of the success of the Web. The simplicity of document publication on the Web allowed scholars to greatly increase the dissemination of their work.

In the light of these characteristics, we are focusing on three main issues for scholarly hypermedia research: the focus on metadata, the study of new interaction modalities and the design of new tools. Foremost, *metadata* is a crucial part of intelligent information handling, especially in scholar work [9]. In the Web, hyperlinks can be considered as annotations on the text, that provide an explicit structure. Metadata becomes all the more important as we deal with documents of other natures (images, video, sound). Moreover, the *increased interaction possibilities* allow to envision innovative uses of hypermedia, but we have to learn how to conceive and use them properly, thus provide tools to experiment with and validate new ideas. Eventually, there is a *need for available and accessible tools and technology*, taking inspiration from the success of the Web because 1/ it used simple principles and was based on standards or common practices; 2/ it did not aim at perfection (not implementing for instance a full hypertext vision with complete bidirectional links and rights management); 3/ tools and technology for creating and consulting were widely available (a simple text editor and an open-source web browser, plus internet)

The issues underlined above are all the more relevant as advances in technology now allow us to manipulate in the digital world new media types such as audiovisual documents [10, 7, 27], virtual reality artifacts, augmented reality artifacts [24], etc.

In this article, we will focus on audiovisual (AV) documents, and on the possibilities offered by hypermedia applied to video that we call hypervideo. The first section proposes an overview of scholar uses of audiovisual material, and expresses requirements for scholar-adapted audiovisual information systems. The next section defines our notion of hypervideo, precising how it articulates with existing work. Thereafter, section 4 presents Advene, a working model and prototype framework dedicated to the production and experimentation of hypervideos. The last section discusses the positioning of our approach with respect to existing models and experimentations.

2. AUDIOVISUAL INFORMATION SYSTEMS FOR SCHOLARS

Scholar's use of textual documents, and by extension hypertextual documents, is common and the Web has provided a platform for a better interaction with these, allowing the distribution of research results and articles, the building of databases such as DBLP and the creation of powerful search engines – dedicated or generic. Interaction with other types of material, especially temporal documents such as video or audio, has however been limited to simple consultation, mainly due to the lack of available tools and technological capabilities. It is now possible to envision new uses of audiovisual documents.

Scholar's use of audiovisual documents covers many aspects, as detailed in [2]: AV documents as pedagogical tools (for teachers and speakers in conferences [3]), as a testimony of the past [31] (for historians), as a work of art (for art critics), as personal notes [16] (for anthropologists, sociologists, etc.), as communication acts (for semioticians), etc.

Working with video in a scholar context involves at different degrees a/ audiovisual information retrieval and management; b/ interactive use of the video (at least playback and navigation facilities); c/ active reading abilities (document annotation for facilitating further use); d/ integration of AV material in new documents/publications.

As stated in [1], AudioVisual Information Systems (AVIS) aim at providing means to search, retrieve and use audiovisual documents. Some of them are already in wide use (Virage [33], used by CNN for instance), and research is very active [30, 14], facing the need to handle vast quantities of video material. AVIS processing can be analysed using the following steps: indexing/retrieval (e.g. keywords-based video fragment search), results selection (e.g. selecting most relevant fragments using key-frame and transcriptions), results exploitation. Exploitation depends on the needs of the user but most of the time, it only consists in visualising the selected videos or fragments.

AVIS are much needed in the context of scholar use of AV documents, and must offer a way to accommodate existing practices as well as to promote new interaction modalities. Indeed, current scholar practices often rely on simple video bookmark systems, when not limiting themselves to simple text processing with manual video player control. AV document creation (e.g. for public conference presentation) relies on classical editing tools.

In order to make more advanced uses possible, metadata is essential, providing the basis for all of navigation, active reading (annotation [19]) and further document generation. Metadata is all the more important that AV documents do not feature the same pre-existing, objective indexes that words and letters constitute for textual documents. Moreover, cultural and legal limitations, that are more stringent and strictly enforced on AV documents than on textual documents, make it difficult to share both video and video analyses (for instance a new video with annotations embedded as captions). There is thus a strong need for separation of analysis information (metadata) from the actual documents, so that it does not suffer the same diffusion restrictions.

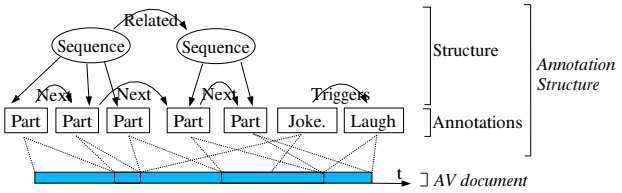

Figure 1: Annotated Audiovisual Document

We can then underline the following requirements for building scholar-adapted AVIS. First, they must support active reading of AV documents, through accessible interfaces, able to accommodate the variety of needs in annotation and visualisation of information. Second, as scholarly annotations, according to [9], need a formalized and pre-defined structure while personal annotations benefit from unconstrained and emergent discovery, tools should help bridge the gap between both, and provide schema evolution for metadata. Besides, AVIS need to overcome cultural and legal limitations, allowing scholars to distribute/publish their analyses independently from the audiovisual documents. Eventually, the existing corpus of information (current video archives) cannot be ignored. AVIS must integrate into existing infrastructures and be able to process existing material without requiring a complete re-processing of the documents (as is the case in the embedding of information as captions in videos).

New usages will foster the development of new interaction modalities with AV documents. The notion of hypervideo, that we present in the next section, offers a framework to analyse and design video-based hypermedia documents.

3. HYPERVIDEOS

Developing new interaction practices in the audiovisual domain requires to build upon existing practices, whose analysis can provide guidelines and directions for the construction of new methods and tools. We propose here a definition of hypervideo that focuses on metadata and the various ways it can be exploited.

We first define (see figure 1) an *Annotated Audiovisual Document* (AAD) as an audiovisual document (AVD) associated with a metadata annotation structure (AS), that features spatiotemporal relationships with the document (i.e. some elements from the AS are linked to spatiotemporal fragments of the video). Elements from the Annotation Structure can also be themselves inter-related (for instance, the *Sequence*s that group multiple *Parts*). No specific constraint is applied on the element's contents, which can be of any type.

A *view of an AAD* can then be defined as a "way of visualising" the annotated audiovisual document, using information extracted from the AV document and information

from the annotation structure. Defined in this very generic way, views can be considered through two main characteristics: 1/ the origin of information for view rendering (AVD and/or AS); 2/ the existence of an access to the AV document temporality. The origin of information for a view can be exclusively the AV document (for instance, its simple visualisation), exclusively the AS (for instance, a generated plain table of contents of the video built from the AS), or the view can mix a number of pieces of information from both sources. Access to AV document temporality means the ability to visualise the AV document and control its playing.

We can then define a *hypervideo* as a view of the AAD that on the one hand uses information from both the AVD *and* the AS and on the other hand gives access to the temporality of the AVD. The previous, non-hypervideo, example of a table of contents can be hypervideated by adding links from the generated table of contents to the corresponding fragments of the AVD. A refinement could also include screenshots from the AVD in the table.

The definition of hypervideo we propose here is compatible with and extends previous definitions of hypervideo [26, 27, 10, 7, 29]. By emphasizing the importance of the annotation structure and the variety of renderings that can be obtained by mixing information from the AS and the audiovisual document, it brings a new light on video material handling, and it gives directions for the design of new video-centered hypermedia systems.

On the one hand, the hypervideo notion allows to analyse existing video-based hypermedia, be they from research works cited above, or from widely used systems. DVD menus for instance offer an access to the temporality of the AV document, using information from both the AS (chapter titles) and the video (e.g. screenshots). However, considering a DVD menu as a hypervideo allows to clearly distinguish between the AVD and the AS, and envision new uses for both of them: users could be allowed to enrich the AS, thus defining their own chapters; the chapter list could be used to produce a textual document; etc.

On the other hand, the hypervideo notion can provide a framework for the development of AVIS, as we present in the following section.

4. ADVENE

Advene[1] is a project that aims at empowering users with the possibility to 1/ annotate videos; 2/ visualise the annotations in user-specifiable ways; 3/ share analyses and comments on videos as hypervideos. It is both a data model, and a framework which can be used to experiment with new ways of dealing with annotated audiovisual documents. General principles include sharing annotations and analysis as packages separated from video material (see figure 4) and simplicity of use for widespread diffusion and innovation. We will first describe the Advene data model, then we will focus on the current visualisation functionalities of the Advene prototype[2], and discuss Advene hypervideos in practice.

[1] *Annotate DVds, Exchange on the NEt*, where DV can also stand for Digital Video.

[2] The Advene prototype is freely downloadable from `http://liris.cnrs.fr/advene/`, with an accompanying demonstration package based on the 8-minute video of Ted Nelson's presentation in a panel of HT'03. Examples in this section are based on this package. We encourage the reader to use it as a companion for this article.

4.1 The Advene data model

The Advene data model (see figure 3) provides a placeholder for the various elements of hypervideo : structured annotations, views, queries. It constitutes a generic layer to store, query and visualise data linked to temporal documents. In order to accommodate the various needs of different categories of users, it does not constrain the kind of information that is linked to the document, but only needs ways to access it. Its intent is not to duplicate existing standards and tools (MPEG-7, video annotation tools such as Anvil [16], ELAN, etc.) but to offer an alternative way to exploit their data.

The basic element in the Advene model is the annotation. An *annotation* is a piece of data linked to a part of an audiovisual document called a *fragment*. Annotations can be put in relation with one another by means of n-ary *relations*, which also feature an optional content describing them.

Annotations' and relations' contents and relationships are typed by *annotation types* and *relation types*. An annotation type possesses a name and defines a content-type for its annotations, in the form of a MIME type (`text/plain`, `text/xml`, `application/pdf`, `audio/wav`, etc.). If the type is `text/xml`, it can be more precisely constrained by a structure description (*e.g.* DTD, XML Schema, ...). A relation type also possesses a name and defines a content-type for its instances. In addition, it specifies the number of participating annotations and their respective types.

Annotation types and relation types can be seen as elementary elements of a specific analysis. We define *schemas* as units grouping related annotation types and relation types. A schema thus represents a certain point of view on the analysed document, and may be reused with other AV documents. It contains a number of semantically related annotation-types and relation-types, defining an elementary analysis framework.

For instance, a basic need for exploiting the video capture of a conference speech is to give it a minimum structure. It can be implemented with a schema called *Summary* (figure 2), which defines the following annotation types: *TalkDescription*, *PartOfSpeech* and *AudioComment*. It also defines a *NextPart* relation type that links an instance of *PartOfSpeech* to the next one. The *TalkDescription* type defines a structured content-type that will hold information about the talk: date, place, name of the event, name of the speaker. The *PartOfSpeech* type only contains the title for the given part. To illustrate the versatility of the model, an *AudioComment* type has been added, with a `audio/mp3` mime-type. Using this schema, it is possible to annotate the video of a talk and use the information in various ways that will be described in the following sections.

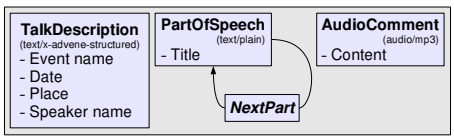

Figure 2: The *Summary* schema

A hypervideo, according to our definition, is a way of visualising an annotated audiovisual document. In order to produce hypervideos, users must be able to define their own

visualisation means. A *view* in the Advene model is a way of visualising elements from the annotation structure and the related video. The model itself does not impose a specific mechanism but the prototype described in the next section (4.2) defines for the moment two different types (static or dynamic) of user-definable views, in addition to hard-coded GUI views.

The Advene project being developed in the context of AVIS, the Advene model also integrates the notion of *query*, that allows to select elements from the model. As with views, the model does not over-specify queries. It simply states that queries return a set of elements from the model, in order to accommodate multiple query models. Details are left to the implementation, which can propose multiple ways to represent and execute queries. The current prototype proposes a simple filtering system, that selects from a set the elements matching a condition. This method covers the basic needs of the current experiments, but this issue is still a work in progress.

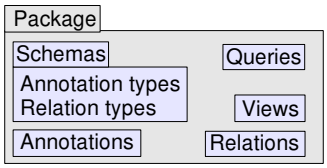

Figure 3: Advene package

The elements of the Advene model (views, queries, schemas, annotation- and relation-types, annotations and relations) form a consistent set: views and queries depend on the schemas, schemas are defined to achieve a specific goal. All these elements are thus gathered in a single unit called *package*, represented figure 3. An Advene package, when accompanied with the corresponding audiovisual medium, contains the necessary material to render hypervideos: the schemas specifying the annotation structure, the annotation structure itself, queries to manipulate it and views to visualise it. A package can be shared by users in order to be simply visualised, or to be further edited, allowing people to build upon other's work.

In order to accommodate reuse needs, two methods are available. A person can simply get a copy of an existing package and modify it to suit her needs. However, she then loses the possible evolutions or improvements of the original package. The other possibility is to reference elements from the original package in the newly created one. Such a reference is called an *import*, and is resolved at load-time. For instance, the *Summary* schema being very generic, it is very likely to be imported for reuse in many packages with the same application domain. Directly importing annotations allows to work on a video reusing previous work.

4.2 Views in the Advene prototype

The Advene prototype is composed of many elements: it implements the Advene model; it proposes a runtime engine able to display various elements related to the annotations; it embeds a web server to serve web documents generated from templates and annotations; it can dynamically control an augmented movie player.

The open-source prototype has been developed in python, which proved to be an ideal language for rapidly and efficiently developing solid, cross-platform applications, already used in similar applications [32]. Portability concerns also had a great influence in the choice of the movie player. We have chosen VLC (`http://www.videolan.org/`), which is an open-source, cross-platform and versatile movie player, able to read video from digital files, DVDs or streams. We developed a python module to control it from our application, and extended its functionalities: it can now for instance display SVG graphics over a video.

The simplicity of the model eases the relationships with existing work from the same domain (information extraction from video). For instance, the Advene import filter can import data from simple timestamped text files, from tools like ELAN or PRAAT, etc. Thanks to the static views, it can also easily export to multiple formats that currently feature CSV spreadsheets, SMIL and SRT (subtitles).

One of the main aims of Advene is to provide ways for annotating video and generating hypervideos, i.e. let users experiment with the visualisation of annotated video documents. The prototype thus implements three types of views: ad-hoc views, User-Time Based Views (UTBV) and Stream-Time Based Views (STBV). Both UTBV and STBV are user-definable, using the Advene runtime engine to generate them, as shown in figure 4.

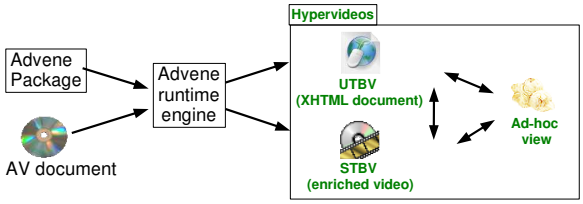

Figure 4: Advene document processing

Ad-hoc views. Ad-hoc views are GUI interfaces to the data. They cannot fully qualify as user-definable views, as they are defined by a source code that only a programmer can modify. However, they can be parameterized to offer some variations in the display of the data. For instance, the transcription view displays annotations of the same type by concatenating their content in a text window. During the movie play, the currently active annotation is highlighted, linking the view to the AV document. This view proposes a number of parameters to customize its appearance: the separator used to concatenate the annotation's contents, a toggle indicating whether to display timestamps or not, etc.

Other available views feature a timeline view displaying the annotations, classified by type, along a timeline; a tree view that allows to navigate into the structure of the package; and a browser view, inspired by NeXtStep file browser, used to interactively evaluate the expressions that allow to address elements in the model.

User-Time Based Views. The first kind of user-definable view, called *User-Time Based Visualisation* (UTBV), may be also called *static view*. It is in fact the definition of a hypertext document, whose temporality is imposed by the user visualising it. The User-Time Based Visualisation presents the available data (from the package and from the audiovisual document) as hypertext documents. The interaction with the user is done through a standard web browser, that connects to a simple web server integrated in the Advene

application.

The natural support for it is (X)HTML and more generally XML, which can be displayed by any web browser and can easily be generated. We have chosen to implement templates using Zope Page Templates (ZPT), a technology issued from the Zope platform (`http://www.zope.org/`). It consists of a template language called TAL (*Template Attribute Language*) coupled with an expression syntax for data access called TALES (*Template Attribute Language Expression Syntax*). The couple TAL/TALES presents a number of interesting properties – simplicity, availability of useful tools and libraries, validity of the XML templates and model abstraction – that are described in [1].

To illustrate, we will just here comment a small excerpt, presented in figure 5, that generates a table of contents with illustrating snapshots. It uses three of the seven processing instructions of TAL: a loop construct (`tal:repeat`), an attribute replacement construct (`tal:attributes`) and a tag content replacement construct (`tal:content`). The loop iterates on all annotations of a given type (which is obtained through a TALES expression), and generates for each element a link offering to play the given sequence represented by a snapshot and the content of the corresponding annotation. Being XML-valid, such code can be edited through a standard XHTML editor that preserves attributes from other namespaces. Indeed, a browser-based WYSIWYG XHTML editor has been integrated in Advene, and it will be augmented with TAL editing facilities.

```
<ul><li tal:repeat="a
  here/annotationTypes/PartOfSpeech/annotations">
 <a tal:attributes="href a/player_url">
  <img alt="" tal:attributes="src a/snapshot_url" />
  <strong tal:content="a/content/data">Name</strong></a>
</li></ul>
```

Figure 5: Valid XML and simple processing instructions for UTBV definition

Figure 6 demonstrates the rendering of static views in a standard web browser. Using the annotation structure, a summary of the speech, displaying the sequences with an illustrating screenshot and their title, allows to access the desired sequence from the webbrowser. Moreover, the Details link displays more information about the sequence (browser window in the right-hand corner), comprising the linked transcription offering an access to more specific parts of the speech.

Following the principles of the Web, various categories of users can define their own views. The most advanced can create original views, but other may create views by imitation of existing ones, copying extracts matching their needs, or simple adaptation. Eventually, instrumented WYSIWYG editors will allow less advanced users to create their own views.

Stream-Time Based Views. In the second kind of user-definable view, *Stream-Time Based Visualisation* (STBV), the temporality of the resulting document is mostly imposed by the audiovisual document. Of course, they also offer some kind of interaction opportunities and the user normally always has the possibility to interrupt playing, but we can imagine kiosk-like approaches where all video controls are deactivated. STBV can be seen as a video augmented with additional capabilities. Of the two main

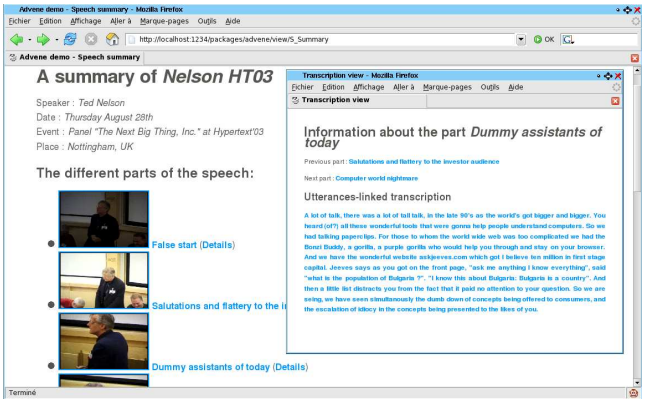

Figure 6: Rendering of a static view

approaches [28] found in dynamic presentation of content (scheduled or event-based), we have chosen to use an event-based model, using the Event-Condition-Action (ECA) paradigm [21] which can be found in many applications such as databases or mail client software for filtering.

The rationale for this choice is that we aim at user accessibility: scheduled views are more programmer-oriented, and the event-based approach is well suited to highly interactive hypermedia [13]. Our use of ECA-rules is not meant as a full-fledged composition language, but as a simple means to achieve goals expressed by the users, using an understandable model. It is however flexible and powerful enough, as we have found out during experimentations. For instance, using both annotations and relations, it is possible to generate on the fly the summary of a speech: annotations define the sequences, and relations define their order. The corresponding view then only has to specify that at the end of a sequence, the player has to go to the beginning of the related sequence.

A STBV is defined by a set of rules. Each rule can be triggered by the occurrence of an event, generated by the application and related to the activation of annotations during the movie play or the changing state of the player (pause, play, etc.). When activated, the rule checks that some user-specified conditions are met, and executes actions accordingly. To implement the actions, the Advene application integrates an enhanced video-player that offers standard and extended capabilities: navigation in the movie stream, screenshot capture, text captioning, display of SVG graphics on the screen, etc. These capabilities are then exposed to the user through a number of possible actions. The currently implemented actions allow to control the standard functionalities of the player (start/stop the player, go to a given position, modify the volume, etc.), control the extended functionalities of the player (take a screenshot, display a text caption, display SVG graphics over the video, etc.), interact with the user (log messages, display dialog popups with video navigation options, etc.), and control the application itself (activate another dynamic view, open an ad-hoc view, open a URL, etc.). The infrastructure of Advene makes it easy to propose and implement new types of actions, thus offering an excellent playground for testing new hypervideo interaction modalities.

For instance, using the example presented in section 4.1, it is possible to design a view that displays the sequence title

and proposes to navigate to the previous or next sequence. Figure 7 presents the GUI that allows the user to define the adequate ruleset: when the beginning of an annotation occurs, and if the type of the annotation is *PartOfSpeech*, then caption the annotation with the sequence title (extracted from the annotation's content) and, using the relations indicating the next and previous sequence, display a popup dialog proposing to the user to go to the previous or next part. The screenshot in figure 8 (section 4.3) presents the result of the application of this view: the navigation popup in the lower left-hand corner proposes to navigate to the next and previous sequences, and the current sequence title is displayed as a caption over the video. Bringing this kind of new interaction possibilities is only a matter of defining a rule.

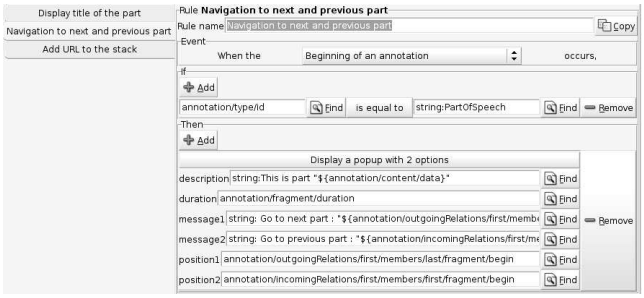

Figure 7: One of the three rules of a Stream-Time Based View

No effort is made by the application to ensure the global validity of a set of rules, except for basic safeguards. A great deal of flexibility is given to the user, which may allow the definition of conflicting rules, for the sake of versatility.

As with User-Time Based Views, the principle of the STBV allow various categories of users to define new views: advanced users can create original ones, or imitate existing ones. With the growing use of Advene, we hope that a number of generic views will emerge, that can be proposed as-is for basic users, or generated through assistants that will guide the user step-by-step. We hope that this simple framework for building new dynamic visualisation and interaction possibilities will allow users with various interests to experiment with innovative ways of presenting information.

4.3 Hypervideo in practice

The Advene prototype addresses the issues of scholarly hypermedia research identified in the introduction. First, Advene principles naturally promote the relevance of metadata and their separation from the AV document. Moreover, by providing user-definable view facilities, our prototype provides a new tool to experiment with new interaction modalities with AV documents. Its flexibility also allows us to quickly design new ad-hoc views or components used in dynamic views (such as information popups, history widgets, etc). Figure 8 presents a sample interface, featuring video navigation popups, navigation history and information URLs as interactive facilities.

A first level of active reading is achieved in Advene through the note-taking feature, that proposes to write notes with timestamp delimiters while watching the AV document. Once the notes are complete, they are converted to annotations of a given type. It is possible to indicate that some notes, use-

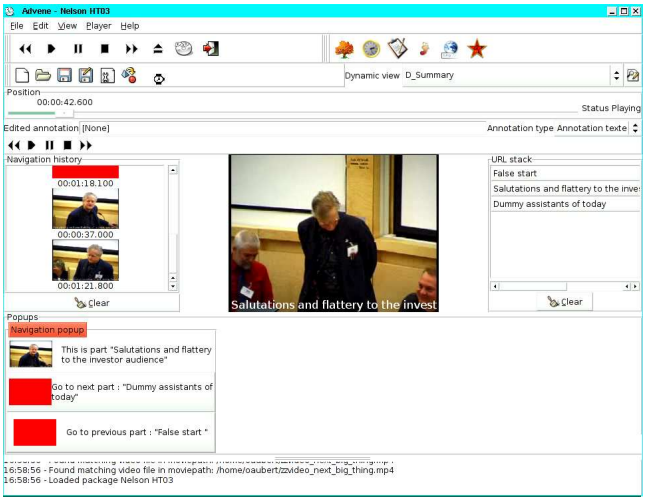

Figure 8: Rendering of a dynamic view

ful as reminders for the user in the annotation process, are to be ignored during the conversion. Figure 9 shows a snapshot of the GUI during the note-taking process: the editing window lies in the lower part of the interface, where squares represent the delimiting timestamps. Red squares indicate that the following text will be ignored in the conversion. The timeline view that overlaps the main window presents the result of the conversion, where edited notes have been converted to annotations of type *PartOfSpeech*.

Annotation assistants can also be used to automatically extract features from the AV document (audio segmentation, shot segmentation, speech recognition, etc) and produce annotations. Refined annotations can then be generated from this raw data (hand-edited or generated by assistants) with a simple drag-and-drop in the timelime window, converting existing annotations to a new type (in order to structure them) or by aligning annotation bounds with other annotations.

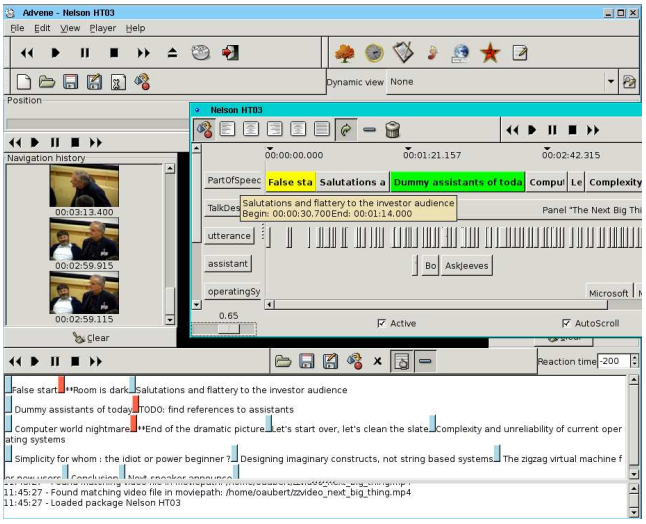

Figure 9: Note-taking with Advene

Both static (UTBV) and dynamic (STBV) user-defined

views can be hypervideos: data can be obtained from both sources (metadata and AV document), and it is always possible to provide the user with an access to the AV document temporality, thanks to the integration of an application-controlled video player and facilities for automatic (event-based) or manual (link-based) triggering of views. For instance, a table of contents (static) view can feature links that launch the video at the appropriate location, with a dynamic view that automatically loads another static view when the end of a part has been reached.

Now that we have a working base, we want to focus on usability issues, such as the extension of a WYSIWYG editor for HTML templates, or the definition of standard dynamic views, so that unexperienced users may also exploit their video documents. These studies will involve humanities scientists, cinema teachers at the academic level and open-source volunteers. With the diffusion of Advene, we expect that common schemas and views will emerge and be stored in reference packages. Such reference packages can be shared (via email) or be made available in package repositories (on web sites), so as to disseminate schemas of videos analyses related to various activities.

5. RELATED WORKS AND DISCUSSION

We have presented in section 3 the notion of hypervideo, and a working prototype aimed at experimentation of new visualisation modalities in section 4. We will now see how the hypervideo model and its Advene instantiation relate to other existing work, and discuss their implications.

5.1 Related works

Hypervideo, and its Advene incarnation, lies at the crossroads of many fields of research: it provides an annotation model, as well as an hypermedia model, and relates to some extent to presentation models and previous work in information retrieval and interface design.

Other annotation models for AV documents have been previously devised, either for specific purposes (Anvil [16]) or with more generic goals in mind (AI-Strata [23], Opales [4], MPEG7 [25]). The hypervideo model can accommodate either annotation model, provided that addressing schemes can be provided. However, the Advene model has been devised with simplicity and visualisation issues in mind, while remaining fairly generic. Care has been taken to ensure a good level of interoperability with other models.

Moreover, the choice of storing all metadata in a single package, separated from the AV document, stems from a number of reasons: first, it allows the enrichment of existing AV corpuses without having to re-encode all material. Second, it does not impose a canonical view on an AV document, but instead allows the use of different points of view (schemas) to analyse the same document. Eventually, it avoids to contaminate the metadata with the diffusion restrictions of the AV document: users are free to share their interpretations of whatever material they have. Whitehead's containment modeling [35] underlines the importance of identifying the storage methods for the various elements of hypertext documents. Other projects with different goals use different methods: the Informedia project [14] stores its data in a database, thus requiring a network access to work with it. Annodex [22], intending to stream metadata with the AV document, embed annotations into it. Embedding metadata in the AV document is appropriate for the diffusion of finalized montages, but it imposes a specific point of view (the one that is embedded) on the AV document, and does not explicit the structure of the accompanying metadata. Moreover, it imposes on the metadata the same diffusion restrictions as those of the AV document, and requires that the AV document is converted in a supported format, while Advene is able to handle any type of AV document (digital video file, streaming data, DVD). This approach is thus complementary of the Advene approach, where Advene can be used to collaboratively work on an AV document, and produce an Annodex file as one of the results. HyStream [3] implements the notion of *continuous metadata*, applying the Open Hypermedia concept to temporal media. Initial assumptions differ from Advene: metadata is not meant to be shared as a document, but rather shared through a central server, allowing to continuously and collaboratively enrich the metadata stream. This approach, used for instance to annotate seminar video captures, prevents disconnected, private operation. However, many similarities with the Advene model can be found, especially in the storage independence of metadata from AV data, and the flexibility of the data model, simply linking data (of any type) to temporal locations.

Advene, as a hypermedia model, can be seen through the filter of the Dexter Reference Model, or of its descendant, the Amsterdam Hypermedia Model, that are meant to provide a generic analysis framework. The Advene model does not have the power and expressiveness of a fully reflexive model like IUHM [20], but it has the virtue of simplicity, its aim being to facilitate the development of new usages.

Table 1 presents how the various elements of the Advene model can be placed in the Dexter Reference Model (DRM) [11] (and its extension, the Amsterdam Hypermedia Model [12]).

DRM	**Advene**
runtime layer	data representation (hypervideo generation)
presentation specification	views
storage layer	package (metadata + structure + views + queries) + video
anchors	expressions + content access plugins + fragments
within component layer	annotation content + video

Table 1: Advene and the Dexter Reference Model

One of the similarities of the Advene model with the DRM is that the within-component layer is not elaborated: the interpretation of component contents and internal structure is left to the tools using the model. Similarly, Advene only provides a temporally-linked information placeholder, able to contain any type of information, with a minimal structure (annotation types and relation types). If the information has to be structured, it will be specified in the annotations' content (through the use of XML Schema or other means).

Related to other hypermedia presentation models, such as SMIL [6] or Cuypers [5], Advene promotes a lax approach to the specification of elements presentation. It allows for instance to specify conflicting rules, in order not to hinder experimentations by the users. Moreover, Advene promotes the use of structured data as far as possible,

and the availability of metadata to the end users, in order to accommodate the variety of visualisation needs. Presentation models such as SMIL provide appropriate rendering targets for annotation data (Advene indeed provides a simple SMIL export static view), but cannot be used as an exchange medium because of the loss of structure information. In other words, SMIL focuses on the presentation of data, while Advene focuses on the data itself (and what can be done with it: sharing, presentation, etc).

Advene does not try to automatically extract information. This task is left to other projects, with the intent of being able to easily integrate external data. For instance, shot detection, caption extraction, speech recognition are fields of research with interesting results that can be integrated. Informedia [14] does this already for a TV corpus (news, TV shows) by providing transcriptions, shot detection, etc.

Eventually, much work has been done around hypervideo visualisation issues [8, 36, 30]) and new interfaces. The Advene project aims at providing a framework to help in the design and experimentation of such new interfaces.

5.2 Discussion

The notion of hypervideo provides a framework to analyse existing systems as well as build new ones. It encompasses previous definitions of hypervideos, while focusing on the relevance of metadata, separated from the AV document, and on the multiplicity of possible visualisations. The Advene project provides a simple but powerful platform to produce hypervideos.

The Advene features match the three characteristic of scholar use of hypertext that we listed in the introduction: linking, interaction and collaboration. First, it offers the possibility to *create links* both dynamic/inside the video (in dynamic views) and with the web (generated static views which may interact with the dynamic views/player). Second, it allows *new interaction modalities*, by offering the ability to define dynamic views where the metadata and the reader both play an active role in the video play. Third, *collaboration is facilitated* by the separation of metadata from the video, which provides sharing and reuse facilities thanks to 1/ the reification of metadata in units (packages), distinct from the audiovisual document, which are storable and sharable; 2/ the ability to import (make reference to) data from other packages.

Considering the importance of metadata, it is crucial for collaboration that it is available to the end-user in a structured, original form, and not processed into a rendering which would be diffcult to further exploit. Advene thus focuses on structured data sharing, accompanied by means of exploiting it. This structured meta-data availability, as well as the ability for users to define their own visualisation means, provides opportunities for user-lead innovation [34].

Among the seven dimensions for considering interactive scholarly communication that Leggett and Shipman propose in [18], Advene is more particularly concerned with three: interaction, indirection and narrative.

Hypervideos produced by Advene indeed provide a very *high level of interaction*, in the sense that not only the reader can intervene during the visualisation process (either on her own initiative, or answering questions and activating links proposed during the playing), but she also can get involved in the creation process, by defining her own views, her own analyses and using the original source document as a base for an extended work.

The clear separation between the AVD and the AS (materialised by the Advene package) also enforces a *high level of indirection*: all visualised hypervideos are the result of the processing of both the video *and* the package. Incidentally, this raises the question of the status of a package as a document, since its content is not meant to be directly consulted, but to produce other document instances that can be consulted. Packages can rather qualify as metadocuments, used to link other documents.

Eventually, both ends of the *narrative dimension* are covered by Advene hypervideos: on the lower end, the annotation structure constitutes raw, reusable information, while on the higher end, views (either static or dynamic) provide a way to build constructed discourses, using the AS as base material.

6. CONCLUSION AND FUTURE WORK

In this paper, we have first presented a number of requirements for enriched scholar use of video material. We have defined the notion of hypervideo, as a framework to analyse existing video-based hypermedia as well as build new ones. The Advene project aims at implementing hypervideos through a generic model and a framework dedicated at experimenting with hypervideos, using user-definable views, static and dynamic. After overviewing related works, we have discussed how Advene answers the requirements for scholar work with AV documents.

The main contribution of the hypervideo notion is to stress the needed separation of concerns between AV data and metadata, related to storage and processing concerns. The Advene project claims that users should be able to specify themselves how they want to display metadata associated to an AV document, and that the metadata should thus always be available in a structured form to the end-users. By proposing an open and available framework, the Advene project aims at fostering the development of new interaction modalities with AV documents, through user-lead innovation.

We are confident that Advene covers the three issues for scholarly hypermedia research we presented in the introduction. Sharing and collaboration being at the root of any scholar work, information should not be hidden in the visualisation material. *Metadata* is essential in the interaction with audiovisual material, and it must be made available to the end-user, so that she can process it any way she wants. Advene packages provide structured and reusable information, independent from the AV document. *New interaction modalities* (both in conceiving and reading) of audiovisual documents are to emerge, and Advene offers opportunities to experiment with hypervideo and find appropriate representations and ways of dealing with hypervideos [18]. Moreover, Advene is an *available and accessible tool* that can accommodate a variety of uses and points of views on AV documents. Its open-source license is intended to foster unhindered innovation, among users such as computer-savvy moviegoers.

Future work includes the conduct of experiments in the ongoing collaboration with a team of researchers using video as a research support for human interaction analysis and the development of new annotation interfaces. With the development of uses, we aim at providing standard schemas and views that can be reused in other contexts. We will also

study package and hypervideo sharing through imports and package repositories. Video active reading, personal annotations, schema design and evolution are also present among our concerns, as well as the study of hypervideo perception and interaction in multiple temporal contexts.

7. REFERENCES

[1] O. Aubert and Y. Prié. From video information retrieval to hypervideo management. In *Corimedia, the international workshop on multidisciplinary image, video, and audio retrieval and mining*, Sherbrooke, Canada, Oct 2004. 10 pp.

[2] G. Auffret and Y. Prié. Managing Full-indexed Audiovisual Documents: a New Perspective for the Humanities. *Computer and the Humanities, special issue on Digital Images*, 33(4):319–344, 1999.

[3] R. Beales, D. Cruickshank, D. D. Roure, N. Gibbins, B. Juby, D. T. Michaelides, and K. R. Page. The pipeline of enrichment: Supporting link creation for continuous media. In *Openness, Structural Awareness, and Adaptivity: International Workshops*, volume 2266 of *LNCS*, pages 47–58, Berlin, Aug 2001.

[4] H. Betaille, M. Nanard, and J. Nanard. OPALES: An Environment for Sharing Knowledge among Experts Working on Multimedia Archives. In *Museum and the Web*, pages 145–154., Seattle, USA, Mar 2001.

[5] S. Bocconi. Automatic presentation generation for scholarly hypermedia. In *1st Int. Workshop on Scholarly Hypertext at HyperText 2003*, Aug 2003.

[6] D. C. Bulterman. Using smil to encode interactive, peer-level multimedia annotations. In *ACM symposium on Document engineering*, pages 32–41, New York, 2003.

[7] T. Chambel, C. Zahn, and M. Finke. Hypervideo design and support for contextualized learning. In *IEEE International Conference on Advanced Learning Technologies (ICALT'04)*, pages 345–349, Joensuu, Finland, august 2004.

[8] W. Ding, G. Marchionini, and D. Soergel. Multimodal surrogates for video browsing. In *Proceedings of the fourth ACM conference on Digital libraries*, pages 85–93, Berkeley, California, United States, 1999.

[9] R. Furuta and E. Urbina. On the characteristics of scholarly annotations. In *Thirteenth ACM conference on Hypertext and hypermedia*, pages 78–79, College Park, Maryland, USA, Jun 2002.

[10] A. Girgensohn, L. Wilcox, F. Shipman, and S. Bly. Designing affordances for the navigation of detail-on-demand hypervideo. In *Proceedings of AVI 2004*, pages 290–297, Gallipoli, Italy, May 2004.

[11] F. Halasz and M. Schwartz. The dexter hypertext reference model. *Communications of the ACM*, 2(37):30–39, ferburary 1994.

[12] L. Hardman, D. C. A. Bulterman, and G. van Rossum. The amsterdam hypermedia model: adding time and context to the dexter model. *Communications of the ACM*, 37(2):50–62, february 1994.

[13] L. Hardman, P. Schmitz, J. van Ossenbruggen, W. R. ten Kate, , and L. Rutledge. The link vs. the event: Activating and deactivating elements in time-based hypermedia. *New Review of Hypermedia and Multimedia*, 6:89–109, 2000.

[14] A. Hauptmann, R. Baron, W. Lin, M. Chen, M. Derthick, M. C. R. Jin, and R. Yan. Video classification and retrieval with the informedia digital video library system. In *Proceedings of the Tenth Text REtrieval Conference*, Gaithersburg, MD, 2002.

[15] U. P. Karadkar, L. Francisco-Revilla, R. Furuta, H. Hsieh, and F. M. S. III. Evolution of the walden's paths authoring tools. *Webnet 2000*, pages 299–304, november 2000.

[16] M. Kipp. ANVIL - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of Eurospeech 2001*, pages 1367–1370, Aaborg, Sep 2001.

[17] D. Kolb. Scholarly hypertext: self-represented complexity. In *Proceedings of the eighth ACM conference on Hypertext*, pages 29–37, Southampton, United Kingdom, 1997.

[18] J. J. Leggett and F. M. Shipman. Directions for hypertext research: Exploring the design space for interactive scholarly communication. In *Proceedings of the fifteenth ACM Conference on Hypertext and Hypermedia*, pages 2–11, 2004.

[19] C. Marshall. Toward an ecology of hypertext annotation. In *Proceedings of ACM Hypertext '98*, pages 40–49, Pittsburgh, PA, june 1998.

[20] M. Nanard, J. Nanard, and P. King. IUHM: a hypermedia-based model for integrating open services, data and metadata. In *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pages 128–137, 2003.

[21] N. W. Paton, editor. *Active Rules in Database Systems*. Springer Verlag, New York, 1999.

[22] S. Pfeiffer, C. Parker, and C. Schremmer. Annodex: a simple architecture to enable hyperlinking, search and retrieval of time-continuous data on the web. In *5th ACM SIGMM International workshop on Multimedia information retrieval*, pages 87–93, 2003.

[23] Y. Prié, A. Mille, and J.-M. Pinon. AI-STRATA: A User-centered Model for Content-based description and Retrieval of Audiovisual Sequences. In L. N. in Computer Science, editor, *First International Conference on Advanced Multimedia Content Processing (AMCP'98)*, volume 1554, pages 328–343, Osaka, Japan, Nov 1998.

[24] L. Romero and N. Correia. Hyperreal: a hypermedia model for mixed reality. In *Proceedings of Hypertext'03*, pages 2–9, Nottingham, UK, Aug 2003.

[25] J. M. M. Sanchez, R. Koenen, and F. Pereira. MPEG-7: The Generic Multimedia Content Description Standard, Part 1. *IEEE Multimedia Journal*, 9(2):78–87, Apr-Jun 2002.

[26] N. Sawhney, D. Balcom, and I. Smith. Authoring and navigating video in space and time. *IEEE Multimedia*, 4(4):30–39, Oct 1997.

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

[28] P. Schmitz. Unifying scheduled time models with interactive event-based timing. Technical report, Microsoft Research, 2000.

[29] T. Schneider and S. W. Smoliar. Description and narrative in hypervideo. In *Proceedings of the 34th*

Hawaii International Conference on System Sciences, volume 4, page 4014, Maui, Hawaii, 2001.

[30] A. Smeaton, H. Lee, and K. M. Donald. Experiences of creating four video library collections with the fischlar system. *International Journal on Digital Libraries*, 4(1):42–44, 2004.

[31] A. Stein, J. Keiper, L. Bezerra, H. Brocks, and U. Thiel. Collaborative research and documentation of european film history: The collate collaboratory. *International Journal of Digital Information Management (JDIM)*, 2:30–39, Mar 2004.

[32] G. van Rossum, J. Jansen, K. S. Mullender, and D. C. A. Bulterman. Cmifed: a presentation environment for portable hypermedia documents. In *First ACM international conference on Multimedia*, pages 183–188, Anaheim, California, USA, 1993.

[33] Virage video systems, 2004. `http://www.virage.com/`.

[34] E. von Hippel. *Democratizing innovation*. MIT Press, Apr 2005. ISBN 0-262-00274-4.

[35] J. Whitehead. Uniform comparison of data models using containment modeling. In *Thirteenth ACM conference on Hypertext and hypermedia*, pages 182–191, College Park, Maryland, USA, june 2002.

[36] C. M. Wildemuth, G. Marchionini, M. Yang, G. Geisler, T. Wilkens, A. Hughes, and R. Gruss. How fast is too fast?: evaluating fast forward surrogates for digital video. In *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries*, pages 221–230, Houston, Texas, 2003.