Interacting with annotated videos

Yannick Prié
Université de Lyon, CNRS
Université Lyon 1, LIRIS,
UMR5205, F-69622, France
yannick.prie@liris.cnrs.fr

CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-246-7/09/04.

Abstract
We present Advene, a project and a prototype dedicated to video active reading, i.e. inscribing annotations on a video document and taking advantage of these annotations for navigating it. Both the video and its annotations are temporal data that need to be presented and interacted with. We present some illustrative interfaces, and discuss some of the challenges related to such a complex activity.

Keywords
Audiovisual annotation, active reading, document time, stream time, video interaction

ACM Classification Keywords
H.5.1 [Multimedia Information Systems] : Video; H.5.2 [User Interfaces] : Interaction styles, Theory and methods; H.5.4 [Hypertext/Hypermedia] : User issues

Introduction: video active reading
Video documents take more and more importance within the computer world. Of course, their most widespread use is related to their basic consumption from beginning to end thanks to basic video players, the main useful commands of which are *play* and *pause*. Nevertheless, some video-related activities go beyond such basic consumption and need *deep manipulation* of videos. Amongst those activities are

montage (creation of a video from several video and audio tracks) and active reading.

Active reading is a so-called "knowledge-work" activity that consists in taking notes on documents while reading them, so as to be able to produce new documents. Such notes are called annotations; they usually feature an anchor (link to a fragment of the document) and content. In audiovisual active reading, annotations anchors usually are temporal fragments of the stream (defined with two time codes), while their content is most often textual. New documents resulting from active reading sessions fail into the general category of hypervideos. Video active readers include ergonomists, linguistic or gesture analysts that annotate video captures of activities for analyze; film enthusiasts or critics who annotate movies so as to publish and share their analyses as hypervideos.

The Advene project and prototype

The Advene project (Annotate Digital Video, Exchange on the Net, http://advene.org) aims since 2002 at studying the emergence of hypervideos – as new hypermedia forms created upon audiovisual documents – and at designing systems that allow creating, building and sharing hypervideos [2]. It features a data model [3] that defines packages as container of annotations, relations between annotations, description schemes, and views definitions related to one video. Views can be either static (web-based, e.g. a table of contents in a browser) or dynamic (player-based, e.g. subtitling of the movies during playing). The set of views related to a video compose a hypervideo that can be shared by sharing the package (annotation structure, view definitions) it is generated from.

The Advene prototype [1] is a free, python-based, multi-platform, open source application that includes a video player, and offers active reading facilities, implementing the Advene data model. It allows the user to annotate video documents and to define both static and dynamic views so as to express her personal analysis that can be further shared and refined. It also features several GUIs that are used for annotating the video stream and managing its annotations.

Two kinds of temporal data

There are two kinds of temporal data that need to be presented and interacted with during video active reading. First, the audiovisual document itself is natively temporal, which means that it mandatory needs a player to be played. Two temporalities are related to such documents: the *stream temporality* corresponds to the stream being played at each instant at a certain rate; the *document temporality* comes from the duration of the document and signifies that is has a beginning, an end, and a certain length. Second, the annotation structure is composed of interrelated annotations. Each annotation is anchored in the temporal stream, inheriting its temporality from its anchor fragment. A set of annotations constitutes temporal data over the video document.

Interacting with video to annotate it and with annotations to access video

We have designed several GUIs so as to be able to carry active reading activity. Some of them are presented on figure 1. The most prominent interfaces are a video player to (dis)play audiovisual information, and a classical timeline that presents document time over annotations categories. Trough the representation of the document temporality, it offers navigation within

the document; presenting the annotation structure it allows its creation and management through various interaction modes.

The treeview is a classical hierarchical non-temporal presentation of the information contained in the opened package.

The note-taking view allows the user to take note over the video stream by inserting time stamps inside her text, making it an interactive access to the video. Time-stamped text can further be converted to annotations.

The active bookmark view allows taking snapshots of the video stream and further describing them. It is then possible to combine snapshots (time stamps) into annotations (temporal fragments).

The note taking view adds temporal markers (here represented as images) to a sequence of characters, thus defining annotations fragment and content.

The Active bookmarks view offers direct manipulation of bookmarks and annotations as pairs of bookmarks.

The annotation structure is presented in the timeline view. Abstract annotation types give vertical organization. Relations (here *dialog* > *image*) permit to express more semantics that just plain annotations. The spatialization of the annotations temporalities allows simple manipulation of complex data (e.g. setting the end time of an annotation as the end time of another).

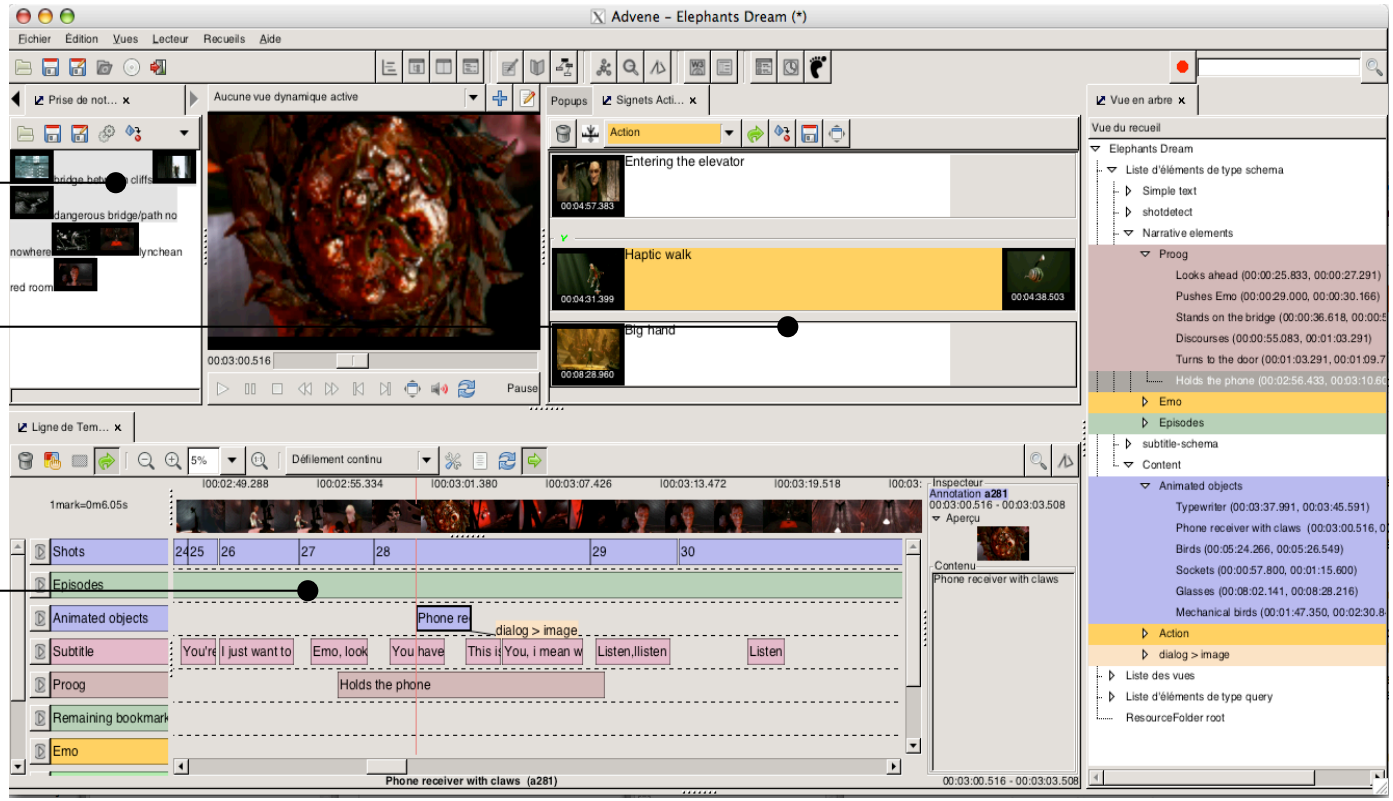

Figure 1. The Advene main interface with several *ad hoc* views: timeline, treeview, note-taking, active bookmarks.

Interaction challenges

The interfaces we have rapidly presented illustrate several challenges that are related to interacting with video annotations.

First challenge consists in being able to *interact with an audiovisual annotation structure* (structure = annotations + relations + description schemes), which is data that on the one side is inherently *temporal* (from its inscription into the stream through fragments) and on the other side is *atemporal*, as a knowledge structure. A table of contents of a film, or the description of the diegetic relationships between characters in the movie can be considered both as atemporal and temporal, whether they allow the playing of the stream, are presented in a temporal or a sequential manner, etc. An annotation can be rightly or erroneously considered both as belonging to a static structure interpretable in the absence of the movie, and as being tightly connected to it, so that it cannot be interpreted without actually playing the related fragment.

As such playing takes time, a sub-challenge relates to *static representation of a temporal stream and its fragments* in the case of video. Video being visually pregnant media that can impress user, fragments can be presented as atemporal structures of excerpt images that can act for the user as mental indexes for her memory of the video stream. Beside classical use of one (begin) or two (begin + end) images for representing fragments, there is plenty of space for new ways of statically representing them, the document they belong to and their relations [5].

Our second general challenge concerns *the management of the temporality mix in the activity of active reading of a temporal stream*. Shortly presented, GUIs for video active reading have to take into account the *stream temporality* (and the associated fascination for the reader) and the *inscription temporality* (temporality of the annotation activity and of all its subtasks, e.g. create an annotation, adjust its begin or end time, type in its content, etc.). The major problem being that annotations are created as a consequence of the playing of a video, but that this very playing becomes annoying as soon as the annotation has to be inscribed, because the task of creating the annotation and the task of going on watching the video collapse. Important issues related to this second challenge relate to loss of attention and recovery, controlled task interruption and resumption [4].

References

[1] Aubert, O. and Prié, Y. Advene: active reading through hypervideo. *Proc. ACM* Hypertext'05 *(*2005), 235–244.

[2] Aubert, O., Champin, P.-A., Prié, Y. and Richard, B. Canonical processes in active reading and hypervideo production. *Multimedia Systems*, 14, 6 (2008), 427–433.

[3] Champin, P.A. and Prié, Y. Models for sustaining emergence of practices for hypervideo. *SADPI'2007* (2007), 59-68.

[4] Iqbal, S. T., and Horvitz, E. Disruption and recovery of computing tasks: field study, analysis, and directions. CHI '07 (2007), 677–686.

[5] Money, A.-G. and Agius, H. Video summarisation: A conceptual framework and survey of the state of the art. *Journal of Visual Communication and Image Representation*, 19, 2 (2008), 121-143.