Structured Media for Authoring Multimedia Documents

TIEN TRAN_THUONG and CECILE ROISIN
Opéra Project, INRIA Rhône-Alpes, Zirst - 655 avenue de l'Europe - 38330 Montbonnot Saint Martin
(tien.tran_thuong@inrialpes.fr, cecile.roisin@inrialpes.fr)

Abstract

This paper proposes a new way for authoring multimedia documents. It uses the concept of structured media that allows to deeper access into media objects. An experiment of structuring video in our authoring and presentation environment for multimedia documents is described. It allows the author to interactively specify various audiovisual structures. A video description and structuration model is used for the composition of video elements (character, shot, scene, etc.) with other media objects (text, sound, image, etc). Such a composition allows to easily realize attractive multimedia presentations where media stuffs can be synchronized in rich and various ways (spatio-temporal synchronization, temporal links, etc.).

1. Motivation

Multimedia document refers to the integration of several media (such as text, image, audio, video and animation) into a document. So a multimedia document model integrates media description models together with temporal and spatial models [1]. The media description model allows to define, to locate, to describe and to group the media that will be used to compose a multimedia document. The temporal and spatial models enable the author to organize media objects in time and space. Numerous approaches have been proposed for multimedia document modeling, including the absolute time axis, the time point temporal model, the interval temporal model, the region model, etc [2]. One of the most interesting result is the emergence of the SMIL standard [4]. However, the media description model mainly consists in declaring the set of used media with their intrinsic spatial and temporal properties. As a consequence, a user can only express coarse-grained relationships (both temporal and spatial relationships) between the different media. But it is worth noting that most media have a rich content information such as image, video, long text or included documents such as HTML or SVG. Through using subparts of that content information, the author can compose multimedia documents having more complex and sophisticated presentation scenarios. Examples of such needs are: a character in a video is introduced by displaying a textual description when that character occurs; a word in a text sentence is highlighted when an audio plays out this word; an hyperlink is set on a video object or on a particular region of an image. These scenarios can be easily specified if the authoring system supplies to the author internal media information such as: a start time of the video object in the video sequence for the first scenario; coordinates of the word in the text and time location of word pronunciation in the audio for the second scenario; or coordinates of the video objects and the image regions for the last scenario. In SMIL for instance, it is possible to specify subparts of media in terms of the instant from the beginning of the media. But it is a rather low level and limited way of specification of media subparts.

Using structured media whose information content is described in a higher level will make this content information available for the composition process. A structured media contains not only raw data, but also a hierarchical description of this media content information. Till now, there are many research works for a standard format of content information description. Within it, the most important is the *Mpeg7* standard also known as "Multimedia Content Description Interface" that aims at providing standardized core technologies allowing the description of audiovisual data content in multimedia environments [3]. So standard structured media for multimedia authoring is not a far future and constitutes the first step to allow the editing of more complex multimedia documents.

The work presented here is performed in the context of the development of a multimedia document authoring system called *Madeus* [1]. This prototype allows to compose multimedia documents from a set of texts, images, audios, video, HTML and SVG media. The *Madeus* document model is based on the structured, temporal interval-based and region-based models. In a first stage, we have experimented the structured media approach with video media.

Based on this experience, the discussion in this paper is devoted to the use of structured media to edit complex multimedia documents. The rest of the paper is organized as follows: first an overview about multimedia authoring is discussed in section 2 and our experiment of structuring video media is presented in section 3. We give in section 4 a brief evaluation of this work through its comparison on modeling and editing aspects with existing works. Finally, the current achievements of our work and some perspectives will be given in the last section.

2. Multimedia authoring systems

Multimedia authoring systems can be classified into different levels. The first level comprises media makers such as video editing tools: *Adobe Premiere*, *CineKit*, *Movi2D*, so on. These tools allow to edit video media from a set of video clips, images and/or texts. They include media analyzers tools, so the author can extract subparts from row media. But they are limited to media production tasks and usually use proprietary formats that cannot be used outside them. Multimedia documents, which have more complex presentation scenario and require more flexible presentation services (such as interactions), need the use of higher-level authoring systems. Examples of multimedia document authoring and/or presentation systems are SMIL applications (*QuickTime4.1*, *IE5.5*, *Grins*, *Ezer*, *Fluition*, etc.) [9], *Director* by MacroMedia and our *Madeus* system. Most of these authoring systems use an XML description of the set of used media and of the presentation scenario. Then browsers which implement the temporal and spatial models can present this specification document. The main advantage of this authoring approach is the capability of integrating numerous media into a single document in a rich and flexible way. But as shown above, they have the drawback that the media are considered as black boxes with low level and limited internal selection facilities (such as in SMIL). So right now, if an author wants to use the subpart of a media inside a multimedia document, he has to use an external tool to cut out this media subpart.

Therefore there is a general need for providing authors with tools for describing and using media content inside authoring applications. The availability of standards will allow the exchange of data from media-oriented applications to document authoring and presentation applications. Our work contributes to the modeling [5] and the development of authoring applications that fulfill both media content description needs and multimedia composition.

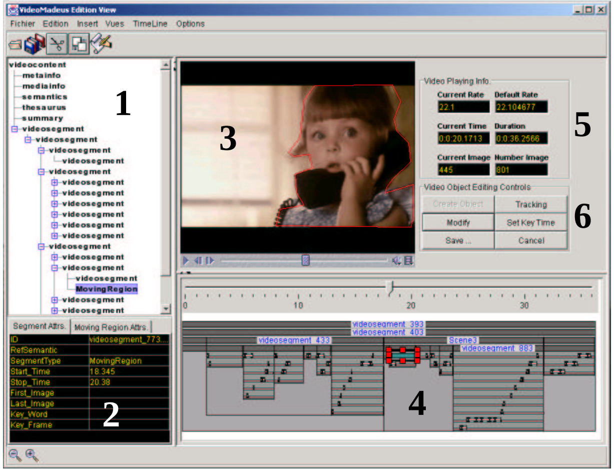
Next section will give more concrete ideas about such authoring systems through our experimental work with the video media.

3. Structured video media experiment

Video is a kind of media, which can carry rich and high-capacity information. Access to internal video components is the key point to build more dynamic and interactive presentations in which video entities can be more fine-grained synchronized with other media. To obtain that result, we have enriched our previous multi-view authoring system *Madeus* [1] with extensions to the document model and with a video content editing view.

We have proposed in [5] a video content modeling for composing multimedia documents. It is based on the hierarchical structure of that media in terms of sequence, scene, shot, transition and object elements as proposed in existing approaches for video indexing [8]. We have also identified other elements that can be relevant for document composition, such as: events (a character goes out a car), moving objects and spatio-temporal relationship elements inside a shot level (two cars move in parallel then one passes in front of the other). Moreover, the semantic part allows not only to group semantically the elements in the video structure but also to describe the actions occurring in the video by including the graph concept, for instance, a footballer kicks a ball. Finally, this model is consistent with our multimedia model so that it is possible to share the same video representation in the different steps of our multimedia authoring system. More precisely, the previous interval-based has been extended with two new elements: the *Sub-Interval* for the time dimension and the *Sub-Region* for the spatial dimension.

In our system, the video content editing environment (figure 1) allows to semi-automatically specify the information within the video media, such as time and spatial internal structure of video presentation using our video description format. The interface presents video content description through several views: the hierarchical structure view (1), the attribute view (2), the video presentation view (3), the timeline view (4), etc.). That provides a simple way for the visualization, the navigation and the modification of the video content description. More concretely, if the author wants to add a video (in the mpeg, avi or mov format) in his document, he simply selects it and the system automatically extracts its basic structure (using a "standard" shot detection algorithm). This first structure is then displayed in the video structure and the timeline views of the video content editing environment. Next the author can adjust and add semantic media content descriptors (such as character, spatial/ personal relation, etc.) which currently cannot be automatically generated by existing content analyzers. For that purpose, some authoring functions are provided: grouping/ungrouping shots, scenes or sequences using the structure view or the timeline view, graphically selecting spatial areas containing objects or characters, attaching key positions and movement functions to these objects using the video presentation view and the attributes panels.



Current Rate	Default Rate
22.1	22.104677

Current Time	Duration
0:0:20.1713	0:0:36.2566

Current Image	Number Image
445	801

ID	videosegment_773...
RefSemantic	
SegmentType	MovingRegion
Start_Time	18.345
Stop_Time	20.38
First_Image	
Last_Image	
Key_Word	
Key_Frame	

Figure 1: The video content description editing view: 1. the video structure view; 2. the attribute view; 3. the video presentation view; 4. the video timeline structure view; 5. the video information view; 6. the video object editing control.

In addition, this video content editing environment has strong relations with other parts of the *Madeus* system allowing to use that video description information when composing *Madeus* documents. Users of *Madeus* can synchronize video elements of a video media with other media objects in both time and space. For instance, in the document displayed in Figure 2, the video object “Little girl phones” of a video segment displayed in the figure 1 has been synchronized with a text media (see the timeline document view). Authors can also apply operations and interactions on elements of the video such as tracking, hyperlink, and erasure [5]. Thus, more complex multimedia documents can be specified while maintaining the declarative approach of XML that allows the use of high-level authoring interfaces like our video content editing.

4. Evaluation

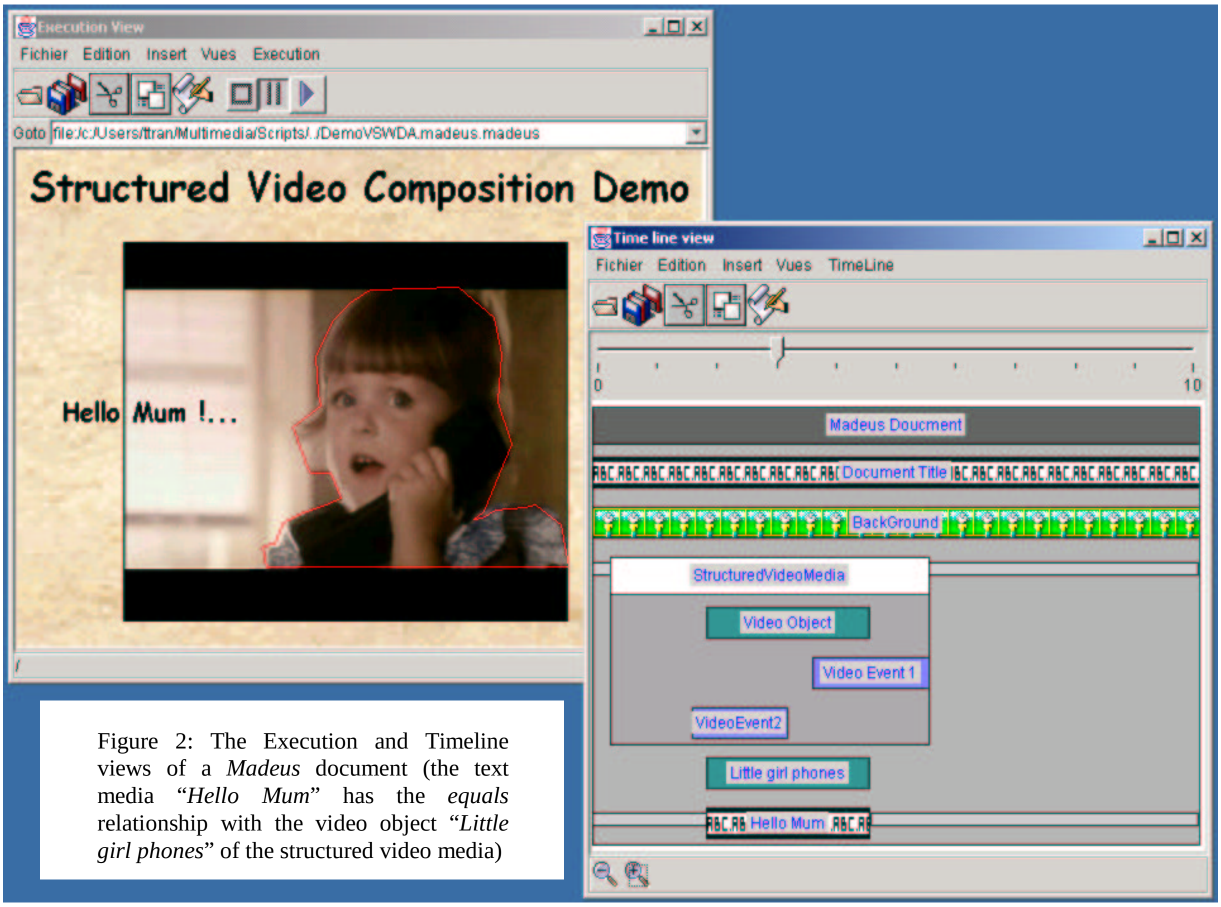
Our proposition provides the support for a deeper access into media content in multimedia document-authoring environments, which until now has treated media content as a black box. In addition, our experimental work with video media has provided a way to implement such system. Notice the media content description model is adapted to compose and render multimedia documents, so it makes little use of metadata descriptions such as Mpeg7 applications mostly devoted for searching, indexing or archiving media content. Indeed, this model is focused on the structural organization of media content that is relevant for multimedia document composition. The media content editing views help the user to create and modify structured media. This environment is similar to the *IBM Mpeg7 Visual Annotation Tool* [6], which is used for authoring audiovisual information description based on the Mpeg7 Standard Multimedia Description Schemes (MDS). However, our tool is open and therefore can integrate automatic media analyzers and generators.

5. Conclusion

In this paper we have proposed the extension of an authoring multimedia environment becoming a more complete application, which can more finely handle media. The experimental development of this application is described. As a positive result of this first experiment, we can edit documents that contain fine-grained synchronizations (in the temporal, spatial and spatio-temporal dimensions) between basic media (text, image, audio and so on) and video elements (such as scene, shot, event, video object. Moreover, we have proposed a semi-automatic tool integrated in the system that analyzes, generates and allows the editing of the content description of video media This result has encouraged us to continue to structure other media. In a next step, we will investigate the same approach for handling audio and text media that allow to compose complex documents as Karaoke document type, in which video, audio and text can have between them fine-grained synchronizations. In addition, the emergence of standard formats for audiovisual information description (Mpeg7) will allow the integration in the media flow of both the content and the description. Therefore multimedia applications will more easily treat these media objects.

References

1. L. Villard, C. Roisin, N. Layaïda, *A XML-based multimedia document processing model for content adaptation*, Digital Documents and Electronic Publishing (DDEP00), September 2000.
2. S. Boll and W. Klas. -ZYX- *A Semantic Model for Multimedia Documents and Presentations*. In Proceedings of the 8th IFIP Conference on Data Semantics, January 1999.
3. *MPEG-7 Documents*, http://www.darmstadt.gmd.de/mobile/MPEG7/Documents.html.
4. *SMIL-Synchronized Multimedia Integration Language*, http://www.w3.org/AudioVideo/.
5. C. Roisin, T. Tran_Thuong, L. Villard, *A Proposal for Video Modeling for Composing Multimedia Documents*, MultiMedia Modeling (MMM2000), Nagano, Japon, November 13-15, 2000.
6. B. Lugeon and J. R. Smith, *MPEG-7 Visual Authoring Tool*, IBM T. J. Watson Research Center, http://www.alphaworks.ibm.com/tech/mpeg-7.
7. R. Hjelsvold, S. Vdaygiri and Y. Léauté, *Web-based Personalization and Management of Interactive Video*, WWW10, May 1-5, 2001 Hong Kong.
8. J. Hunter, J. Newmarch, *An Indexing, Browsing, Search and Retrieval System for Audiovisual Libraries*, Third European Conference on Research and Advanced Technology for Digital Libraries, ECDL '99, 22-24 September, Paris.
9. *SMIL players and authoring tools*, http://www.w3.org/AudioVideo/#SMIL



Execution View
Fichier Edition Insert Vues Execution
Goto file:/c:/Users/ttran/Multimedia/Scripts/../DemoVSWDA.madeus.madeus
Structured Video Composition Demo
Hello Mum !...
Time line view
Fichier Edition Insert Vues TimeLine
0
10
Madeus Doucment
Document Title
BackGround
StructuredVideoMedia
Video Object
Video Event 1
VideoEvent2
Little girl phones
Hello Mum

Figure 2: The Execution and Timeline views of a *Madeus* document (the text media *"Hello Mum"* has the *equals* relationship with the video object *"Little girl phones"* of the structured video media)