SUMMARIZATION OF SCALABLE MULTIMEDIA DOCUMENTS

Benoît Pellan and Cyril Concolato

Institut TELECOM; Télécom ParisTech; CNRS LTCI

ABSTRACT

The summarization of a multimedia document is a challenge that requires the summarization of media elements combined into a document but also relies on an appropriate adaptation of its presentation. In this paper, we present a scalable multimedia model that structures the multimedia scene into incremental *Spatial*, *Temporal* and *Interactive* layers and progressively provides presentation details. Our proposal consists in summarizing such scalable multimedia documents based on three adaptation parameters: a targeted level of expertise, a preferred duration and a level of expectation for extended information. Our approach has been technically validated on PowerPoint-like documents using a generic MPEG-21-based adaptation framework.

1. INTRODUCTION

Content summarization intends to lessen the quantity of data a media conveys in order to minimize its perceived complexity. By extension, document summarization can be defined as a transformation process that consists in reducing the amount of information of a multimedia document and ends up with a simplified presentation of the initial content.

The summarization of media elements has been largely tackled through the use of MPEG-7 descriptions that enable content filtering algorithms based on hierarchical or sequential summaries [1]. For instance, video summarization can be performed according to user queries (personalization) based on relevance metrics expressed using MPEG-7 descriptions [2]. Additionally, media summarization can be extended to the selection of various media modalities as introduced by the *InfoPyramid* framework [3]. Furthermore, the combined use of MPEG-7 and MPEG-21 DIA [4] descriptions also enables the summarization of spatial and temporal properties of scalable media as illustrated by the *Semantic Adaptation Framework* [5]. However, we believe that the summarization of each individual media element of a document is not sufficient to guaranty the generation of a simplified multimedia presentation that is possibly composed of several audio sequences, videos, images but also graphical elements (e.g. polygons, curves) that are styled according to presentation semantics. Therefore, we propose to focus on the summarization of the multimedia scenes (presentation) which describes the spatial and interactive composition of media elements over time.

The approach presented in this paper relies on a document model previously introduced in [6] that divides the multimedia scene description (e.g. in MPEG-4 BIFS or in W3C SVG/SMIL formats) into three scalable dimensions: *Spatial*, *Temporal* and *Interactive (STI)*. The scalability layers of each *STI* axis are incremental and can be selected to generate multiple presentations of the same multimedia document. The media elements and the incremental *STI* properties of the multimedia presentation constitute a scalable multimedia document. Existing approaches have tackled the summarization of multimedia documents by defining an abstraction layer over the multimedia functionality of the targeted presentation formats. For instance, a set of *relations* between media elements is defined in [7] to enable an automated summarization algorithm of the presentation. Our approach does not define such an abstraction layer that often narrows multimedia functionality to a common set, but it structures the multimedia presentation into scalability layers that have progressive requirements in terms of quantity of information. Moreover, we believe that an optimal summarization of a multimedia document requires a clear control from the content authors over the summarized presentation alternatives, even more on structured documents than on individual media. Indeed, the usability [8] of a multimedia document is strongly related to its ergonomic aspects, legibility or accessibility properties. Therefore, specific graphical charters and designer decisions are required to properly compose summarized media elements into a summarized presentation. Thus, we somehow reinforce the *Multimedia Scene Semantic Adaptation* framework defined in [9] by replacing the *Semantic Information Declaration* descriptors; that provide explicit author-driven adaptation guidelines; with scalability layers identifiers that describe the targeted summarization level for each authored *STI* layer: level of expertise (Spatial), preferred duration (Temporal) and need for deeper knowledge (Interactive).

This paper is organized as follows: Section 2 recapitulates the underlying scalable document model of our system (*Scalable MSTI*). Section 3 describes the summarization principles of our approach and Section 4 discusses experiments made with the MPEG-21 framework. Finally, Section 5 concludes this paper.

2. SCALABLE MULTIMEDIA MODEL

The *Scalable MSTI* model defined in [6] clearly separates media elements from the multimedia presentation. In practice, a *scalable MSTI* document can be defined as a logical structure referencing media elements (*Media* description) and on which a set of transformations (*Spatial*, *Temporal* and *Interactive* scene update commands) can be applied to generate a multimedia presentation that addresses various usage environments. The inherent scalability features of such a document are derived from the core components of the *Scalable MSTI* model (*Spatial*, *Temporal* and *Interactive* descriptions) that are further divided into incremental layers as illustrated in Fig. 1. As a consequence, three scalable axes are defined and can be mapped to any progressive adaptation parameter to define an adaptation axis. In this section, we briefly review the *Spatial*, *Temporal* and *Interactive* scalability axes and illustrate their incremental properties on typical adaptation scenarios.

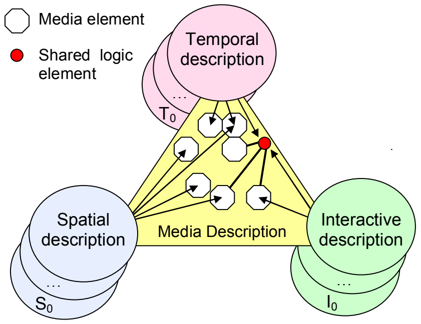


Fig. 1. A scalable multimedia document.

2.1. Spatial scalability axis

The *Spatial* description defines the layout for all media elements that are part of the document. It includes the position of the media elements, the media size, and the visual activation of media elements. A typical adaptation parameter which can be mapped onto the *Spatial* axis is the targeted screen resolution. In that case, the layered description of the spatial properties is ordered to successively address progressive scene size (e.g. S_0:320x240, S_1:400x240, S_2:480x270).

2.2. Temporal scalability axis

The *Temporal* description defines the timing of the multimedia presentation as a whole but also the timing of each individual media element. It usually includes local timer definitions, sequential or parallel behavior and the configuration of timed animations. A typical adaptation parameter which can be mapped onto the *Temporal* axis is the available processing power (or battery state). In that case, the *Temporal* layers are ordered in terms of processing requirements (e.g. incremental levels of smoothness for an animation).

2.3. Interactive scalability axis

The *Interactive* description adds interactive aspects to the multimedia document and defines the behavior associated with interactions. It usually specifies navigation schemes but also keyboard events and mouse actions such as hovering effects. A typical adaptation parameter which can be mapped onto the *Interactive* axis is playback memory requirements. Indeed, the *Interactive* description can be divided into layers by progressively providing interactivity functions to the user. In that case, media elements accessible through auxiliary services are prioritized according to their memory requirements when loaded.

3. MULTIMEDIA SCENE SUMMARIZATION

The content summarization approach introduced in this paper basically consists in reversing the summarization problem from an authoring point of view. Indeed, we propose to tackle the need for a simplified multimedia presentation by authoring a *Base* document whose perceived complexity is as low as possible according to the content creator's criteria and by providing progressive extensions that enrich the semantic information of the presentation. Therefore, the term 'scalable' and 'scalability' should be understood as in the SVC standard [11] and "refer to the removal of [video] bit-stream in order to adapt it to the various needs or preferences of end users". As a consequence, content creators who envision summarization scenarios organize the *Spatial*, *Temporal* and *Interactive (STI)* axes of their *Scalable MSTI* documents in a progressive manner so as to enable the straightforward summarization process that is illustrated in Fig. 2. In this section, we present three adaptation (summarization) parameters based on content-specific interests [12] that can be mapped onto the scalability axes of our document model to generate ready-to-be-summarized presentations. These principles are illustrated on a PowerPoint-like presentation example that can be found at the following address: http://www.tsi.enst.fr/mm/MSTI/SumScalablePres.html

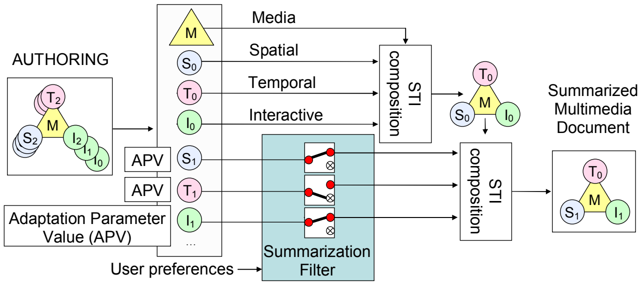


Fig. 2. Scalable scene summarization process.

3.1. Region Of Interest (ROI)

Document summarization may require reducing the quantity of multimedia data conveyed by the document at a given time to enable “at a glance” presentations. For instance, the summarization of the spatial properties of a multimedia scene can be performed by reducing the number of media elements, by modifying the nature of media elements (e.g. through transmoding techniques) or by simplifying the layout of the presentation. The efficiency of such a summarization process heavily relies on an appropriate evaluation of the importance of media elements (e.g. key images but also styled rectangles that may associate an image and its caption).

In the *Scalable MSTI* model, we propose to control the level of details of the multimedia presentation at a specific time through *Spatial* layers. Hence, each layer of the *Spatial* axis define a set of new *Regions Of Interest (ROI)* that progressively complete the presentation with additional details while maintaining the same scene size. The position and size of a *ROI* can be modified from one *Spatial* layer to the next and media elements that are part of a *ROI* can also be updated. Three *ROI-based* summarized presentations of a scalable multimedia document are illustrated in Fig. 3.

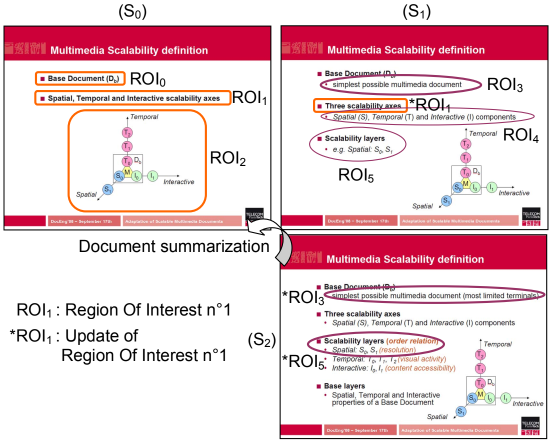

(S0)
(S1)
(S2)
Multimedia Scalability definition
Base Document (Db)
ROI0
Spatial, Temporal and Interactive scalability axes
ROI1
ROI2
Temporal
Spatial
Interactive
T2
T1
T0
Db
M
S0
S1
I0
I1
simplest possible multimedia document
ROI3
Three scalability axes
*ROI1
Spatial (S), Temporal (T) and Interactive (I) components
Scalability layers
e.g. Spatial: S0, S1
ROI4
ROI5
*ROI3
simplest possible multimedia document (most limited terminals)
Scalability layers (order relation)
Spatial: S0, S1 (resolution)
Temporal: T0, T1, T2 (visual activity)
Interactive: I0, I1 (content accessibility)
*ROI5
Base layers
Spatial, Temporal and Interactive properties of a Base Document
DocEng'08 – September 17th
Adaptation of Scalable Multimedia Documents
TELECOM ParisTech
Document summarization
ROI1 : Region Of Interest n°1
*ROI1 : Update of Region Of Interest n°1

Fig. 3. ROI-based document summarization.

3.2. Sequence Of Interest (SOI)

Document summarization may require reducing the quantity of multimedia data proposed by the document over time to shorten the presentation duration. For example, the summarization of the temporal properties of a multimedia presentation can be accomplished by reducing the number of media elements to be sequentially displayed, by shortening the duration of some media elements or by modifying the timing properties of the scene. The optimization of such a summarization process mainly depends on the selection of media elements of the timed summary (e.g. degree of relevance for the presentation understanding).

In the *Scalable MSTI* model, we propose to specify the quantity of multimedia data over a period of time by defining appropriate *Temporal* layers. In our approach, each layer of the *Temporal* axis define new *Sequences Of Interest (SOI)* that progressively extend the presentation with additional media pertinent to the presentation topic as content duration increases. The starting date and duration of a *SOI* can be modified from one *Temporal* layer to the next. Two *SOI-based* summarized presentations of a scalable multimedia document are illustrated in Fig. 4.

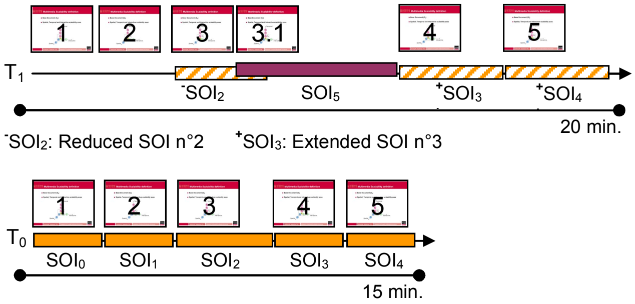
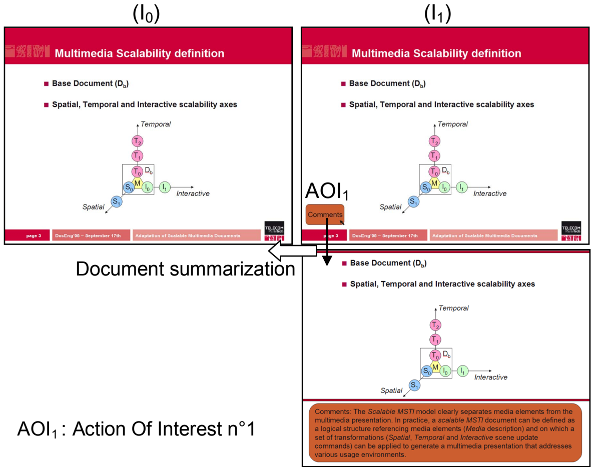


Fig. 4. SOI-based document summarization.

3.3. Action Of Interest (AOI)

Document summarization may require simplifying the visual interface of a document to enhance presentation usability. For instance, the summarization of the interactive properties of a multimedia scene can be performed by reducing the number of user interactive means (e.g. hyperlinks), by reducing the number accessible media elements or by defining limited but simple navigation paradigms. The efficiency of such a summarization process mainly relies on the quality of the media selection that is available because a user interface is only a tool to access media elements. Indeed, a limited access to media elements enable a simpler interface but such improvement of content usability is not satisfactory if essential media elements cannot be accessed.


(I0)
(I1)
Multimedia Scalability definition
Base Document (Db)
Spatial, Temporal and Interactive scalability axes
Temporal
Spatial
Interactive
T2
T1
T0
Db
M
S0
S1
I0
I1
AOI1
Comments
page 3
DocEng'08 – September 17th
Adaptation of Scalable Multimedia Documents
TELECOM ParisTech
Document summarization
AOI1 : Action Of Interest n°1

Comments: The *Scalable MSTI* model clearly separates media elements from the multimedia presentation. In practice, a *scalable MSTI* document can be defined as a logical structure referencing media elements (*Media* description) and on which a set of transformations (*Spatial*, *Temporal* and *Interactive* scene update commands) can be applied to generate a multimedia presentation that addresses various usage environments.

Fig. 5. AOI-based document summarization.

In the *Scalable MSTI* model, the quantity of accessible multimedia data through user actions is determined by the *Interactive* layers. We therefore propose to define new *Actions of Interest (AOI)* in each layer of the *Interactive* axis so that they progressively enable an in-depth understanding to interested users by providing a growing access to auxiliary media elements. The navigation paradigms of an *AOI* can be modified from one *Interactive* layer to the next and media elements that are part of an *AOI* can be updated. Two *AOI-based* summarized presentations of a scalable multimedia document are illustrated in Fig. 5.

4. SUMMARIZATION FRAMEWORK

As illustrated in Fig. 2, in our approach, the summarization process of scalable multimedia documents simply consists in filtering out *Spatial*, *Temporal* and *Interactive (STI)* layers according to the desired level of details. This document summarization can be performed, for example, by a generic MPEG-21 Digital Item Adaptation engine by associating a *generic Bitstream Syntax Description* (*gBSD*) to the scalable multimedia scene. Therefore, the *BSDLink* and more specifically the *AdaptationQoS* descriptions have been used in our experiments to achieve the mapping of *STI* scalable layers on user summarization choices: level of expertise (*Spatial*), acceptable presentation duration (*Temporal*) and need for auxiliary information (*Interactive*). From our experiments, we would like to discuss the following two points.

First, PowerPoint-like presentations are mainly composed of static media (text, graphics elements and images), some animations and navigation mechanisms. The combined use of the summarization axes defined in this paper enables a large set of use cases: short overviews; long but simple presentations; or simple presentations that allow the user to dig into a lot of side information through interactions. However, the simultaneous use of some advanced layers may become awkward: a very detailed presentation that has been minimized to a very short duration may lead to a speedy and therefore unreadable succession of text-heavy slides. For that reason, restrictions on the combined use of some scalability layers using *blocking* or *dependant* layers as introduced in [6] may be defined during authoring to disable such configuration during playback. Second, PowerPoint-like presentations may include video sequences for demonstration purpose for instance. When summarizing such video elements, the combined summarization of the multimedia scene and of video sequences is required. In that case, we could say that we completely reverse the scene adaptation paradigm introduced in [10] by driving the adaptation of media elements from scene-level decisions using MPEG-21-based cross-resource decision-tacking based on global *DIA UCDs* as described in [13].

5. CONCLUSION

In this paper, we propose a global summarization approach for multimedia documents that is focused on presentation properties and that can be combined with media-level summarization techniques. This approach is independent of the presentation format and has been implemented using an MPEG-21-based adaptation framework. Our approach relies on a scalable multimedia model (*Scalable MSTI* model) that is used to generate ready-to-be-summarized presentations by providing progressive details through the definition of *Region Of Interest (ROI)*, *Sequence Of Interest (SOI)* and *Action Of Interest (AOI)*. In the future, we plan to evaluate the subjective impact of the adaptation of a presentation in terms of perceived quality in order to define a quality model that could be used during the content authoring phase.

6. REFERENCES

[1] P. Salembier and J.R. Smith, “MPEG-7 multimedia description schemes”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 748-759, June 2001.

[2] P.M. Fonsesca and F. Pereira, “Automatic video summarization based on MPEG-7 descriptions”, *Signal Processing: Image Communication*, vol. 19, no. 8, pp. 685-699, Sept. 2004.

[3] C-S Li, R. Mohan and J.R. Smith, “Multimedia content description in the InfoPyramid”, in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. IEEE, vol. 6, pp. 3789-3792, May 1998.

[4] ISO/IEC 21000-7:2004, “MPEG-21 Digital Item Adaptation”.

[5] M. Zufferey and H. Kosch, “Semantic Adaptation of Multimedia Content”, in *Proceedings of WIAMIS*, Montreux, Switzerland, April 2005.

[6] B. Pellan and C. Concolato, “Adaptation of Scalable Multimedia Documents”, in *Proceedings of the symposium on Document engineering (DocEng)*. ACM, pp. 32-41, Sept. 2008.

[7] S. Laborie, J. Euzenat and N. Layaïda, “Multimedia Document Summarization based on Semantic Adaptation Framework”, in *Proceedings of the international workshop on Semantically aware document processing and indexing*. ACM Proceeding Series, vol. 259, pp. 87-94, May 2007.

[8] B. Caldwell, M. Cooper, L.G. Reid and G. Vanderheiden, “Web Content Accessibility Guidelines (WCAG) 2.0”. W3C Recommendation, Dec. 2008.

[9] M. Kimiaei-Asadi, J.-C. Dufourd, “Context-aware Semantic Adaptation of Multimedia Presentations”, in *proceedings of ICME*, Amsterdam, Holland, July 2005.

[10] B. Pellan and C. Concolato, “Media-Driven Dynamic Scene Adaptation”, in *Proceedings of WIAMIS*, Santorini, Greece, 2007.

[11] H. Schwarz, D. Marpe, T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard”, *IEEE Trans. on Circuit Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, Sept. 2007.

[12] A. Krapp, “An educational–psychological conceptualisation of interest”, *International Journal for Educational and Vocational Guidance*, vol. 7, no. 1, pp. 5-21, April 2007.

[13] I. Kofler and H. Hellwagner, “MPEG-21-based Cross-Resource Adaptation Decision-Taking”, in *Proceedings of AXMEDIS*, Leeds, UK, pp. 207-214, November 2006.