

# SPATIAL SCENE ADAPTATION IN BROADCAST ENVIRONMENT

*Benoît Pellan<sup>1</sup>, Cyril Concolato<sup>2</sup>*

<sup>1</sup>RTL (EDIRADIO), <sup>2</sup>GET / TELECOM ParisTech CNRS LTCI

## ABSTRACT

The heterogeneity of handheld terminals in terms of screen resolution, processing capabilities or available decoding memory is a challenge for multimedia services that have been tackled by many scene adaptation techniques so far. In broadcast environment, the adaptation intelligence must be transmitted along with the content and may induce critical costs that must be minimized. In this paper, we propose a broadcast-friendly adaptation technique of the spatial layout of multimedia content based on the use of incremental scene updates. The advantages of our approach have been evaluated on a T-DMB digital radio service and compared to other adaptation techniques applicable to broadcasted multimedia services. Experimental results show that fine-grained spatial adaptation on constrained handheld terminals can successfully be achieved through adaptation scene updates with a limited bandwidth overhead.

*Index Terms* — Broadcasted Multimedia Services, Spatial Scene Adaptation, Incremental Scene Updates.

## 1. INTRODUCTION

A multimedia scene describes a collection of media elements, such as video sequences, audio tracks, textual paragraphs or graphic elements, together with a spatial, temporal and interactive organization. Examples of multimedia scenes are advanced web pages or electronic program guides described in HTML, W3C Scalable Vector Graphics (SVG) or MPEG-4 Binary Format for Scenes (BIFS). New generation of mobile devices, like mobile phones, allows users to have access to interactive services based on multimedia scenes. However, the heterogeneity of these devices in terms of screen resolution, processing power or memory is an important issue which currently drives the research for scalable media formats such as SVC [1] or document adaptation techniques [2].

The deployment of broadcast technologies, like DVB-H or T-DMB, allows delivering multimedia scenes to mobile devices. In order to cope with cheap handsets that usually highly constrain the design of broadcasted services, it is critical to design broadcast-friendly adaptation techniques. Hence, traditional adaptation frameworks [3]

that perform content adaptation in a centralized server or in a distributed proxy cannot be applied. Indeed, specific constraints should be considered in broadcast environment. First, since a return channel is not available, or not always available, the adaptation intelligence must be transmitted along with the service. In other words, the content transmitted on the broadcasted channel must contain necessary information for all devices to correctly compose the scene. As a consequence, adaptation techniques [4] relying on XSLT imply unaffordable bandwidth increase. Second, since the adaptation process happens on end user terminal, which is often constrained both in memory and processing, a broadcast-friendly adaptation technique must minimize its computing and memory overhead.

Among all researches done in the scene adaptation field, the adaptation of the spatial layout of media elements raises severe problems. In practice, spatial scene adaptation allows presenting a multimedia scene on different devices with various screen sizes, resolutions and aspect ratios. Existing scene description languages such as HTML, SVG, BIFS or Flash provide some basic spatial adaptation mechanisms: automatic flow/grid layouts for text and/or graphical objects, scalable representation of vector graphics... Additionally, these formats can be extended [5] or semantically annotated [6] [7] to automatically perform enhanced adaptation. However, as described in [8], more author-driven adaptation techniques are required to strictly preserve the legibility of the text content or the graphic charter.

In this paper, we present the state of the art of spatial adaptation techniques and propose a broadcast-friendly approach based on the use of incremental scene description updates. This approach is advantageous for small screen receivers while providing visual enhancements to large screen receivers. Our experiments show the benefits in terms of bandwidth, memory and processing time on typical multimedia services designed for digital radio over T-DMB.

This paper is organized as follows. Section 2 describes the comparison criteria which will guide us in the study of spatial scene adaptation techniques. Section 3 discusses the principles of the state of the art techniques against these criteria while Section 4 develops our proposal. Section 5 presents the results of our experiments and shows the respective benefits of each technique in real cases. Finally, Section 6 concludes this paper and introduces future work.

## 2. SPATIAL SCENE ADAPTATION

### 2.1. General Restrictions

Spatial layout in multimedia documents can be divided into two approaches: explicit and implicit layouts. Explicit layouts assign a 3x2 affine transformation matrix to each media (or group of media) in a global or local coordinate system while implicit layouts gather media into structures according to layout rules (grid, flow, topological). In practice, implicit layout requires more computation than explicit layout (i.e. determining the bounding boxes of neighboring medias), even if it is a more compact representation. For that reason, these features are generally not included in mobile profiles of existing scene description languages. Furthermore, both approaches can be adapted but implicit layout does not guaranty the preservation of author intentions for all display configurations. For both reasons, our study focuses on adaptation techniques tackling with explicit layouts.

As for the generalized spatial scalability of SVC [1], we additionally restrict targeted screen dimensions to incremental values in order to ease the authoring. This restriction orders resolutions and narrows the 2D spatial problem to a one-dimensional problem.

### 2.2. Comparison criteria

In this paper, the objective is to compare adaptation technique results through quantitative criteria such as overheads in terms of processing, memory and bandwidth requirement which are critical when broadcasting adaptable multimedia services on handheld terminals. The *processing* overhead represents the extra computing time (ms) dedicated to handle the adaptation process. The *memory* overhead takes into account the extra data (Kbytes) that need to be loaded on the terminal because of the adaptable feature of a multimedia scene. The *bandwidth* overhead shows the bitrate (Kbits/s) induced by the delivery of adaptation data.

Of course, subjective criteria, such as the adaptation *granularity*, should also be taken into account. Hence, an adaptation technique that addresses any spatial configuration is more interesting than one that focuses on a set of typical screen resolutions. In the first case, the adaptation process is in charge of providing a complete adapted layout to each spatial configuration. In the second case, only a finite number of resolution are targeted and scaling techniques can be additionally used to address others terminal requirements.

Additionally, the flexibility of a spatial adaptation technique should also be noted. This *flexibility* criterion includes the available adaptation features, such as the capability to enrich the multimedia scene with new media elements when the screen size increases, and design efforts required to mix the layout adaptation structures and scene structures (interactivity, timing or animations).

## 3. BROADCAST-RELATED STATE OF THE ART

Given the requirements of broadcast environments and the restrictions exposed in Section 2.1, two spatial adaptation techniques have been selected and evaluated.

### 3.1. Adaptation using alternatives (Switch)

A straightforward approach to provide adaptable scenes in broadcast environment consists in a representation which contains all alternative layouts. Many existing standards provide such feature. The *switch* node used in SVG, SMIL or in BIFS is a typical example. This node groups several media elements and allows the display of only one of them depending on the screen resolution for instance.

This representation has the advantage of being simple in terms of *processing* and allows *flexible* scene changes from one resolution to the next. Unfortunately, it does not scale well with the number of adaptable objects. Indeed, even if an optimized player may discard unnecessary layout alternatives to reduce its footprint and a proper authoring would reuse media, the *bandwidth* and *memory* overheads are proportional to the number of adaptable medias and to the number of adaptation alternatives. Therefore, this approach can only achieve a coarse *granularity*.

### 3.2. Adaptation using Artistic Resizing (ArtResize)

Artistic Resizing is an interpolation technique introduced in [8] targeting the spatial adaptation of user interfaces.

This approach is designed to virtually support any resolution and therefore achieve infinite *granularity*. In terms of *processing*, it requires mathematical computations for each media elements to be adapted but is efficient in terms of *memory* and *bandwidth* since only key values need to be described. The main drawback of this approach is the *flexibility* of linear interpolations. Indeed, appending new media elements requires the definition of new key states of interpolation which may have a significant impact on memory and bandwidth requirements.



Figure 1 - Test scenes with various characteristics

#### 4. SPATIAL ADAPTATION USING SCENE UPDATES

Scene update is a mechanism available in many languages (Flash, BIFS and LAsER) that allows the streaming of timed modifications of a scene. In this paper, we propose to use these updates to achieve efficient spatial adaptation in broadcast scenarios. For a given time, the multimedia service is built by concatenating the scene, designed for the smallest resolution, with adaptation scene updates which either adjust the layout of existing media or insert new media elements. In practice, upon reception of these updates, the terminal applies scene commands in the given order until it detects that the next update resolution will exceed its capabilities. The detection and filtering mechanism can be achieved by inserting specific updates between adaptation scene updates to signal the spatial characteristics required to process the related data. Such signaling, illustrated in Figure 2, is equivalent to the use of specific NAL unit type in the upcoming MPEG-4 SVC standard [1].

From a content provider point of view, layout updates can be automatically generated using interpolation techniques and a predictive coding of updated spatial values can be applied. Moreover, adaptation scene updates does not need to be provided to the terminal at the same time as the base scene. In that case, a proper progressive adaptation of the multimedia scene will be displayed on the user terminal.



Figure 2 - Spatial scene updates organization

The strength of our approach is that it has an excellent *flexibility* since appending new media is natively supported and no particular adaptation structure is required. Moreover, adaptation scene updates do not have any *memory* overhead since only relevant media elements are loaded. Fine adaptation *granularity* can be achieved but the progressive (and possibly overwriting) layout scene updates have an impact on the *processing* required for high resolution terminals. Indeed, numerous writing accesses to the scene tree decrease overall adaptation efficiency. Furthermore, even though predictive coding is used, the consumed *bandwidth* still depends on the number of targeted screen resolutions. Therefore, the granularity of adaptation scene updates has to be selected according to usage scenarios.

#### 5. EXPERIMENTS AND RESULTS

In this section, we present experiments that have been conducted on adaptable T-DMB multimedia radio services. The quantitative results depicted in Section 5.2 compare adaptation techniques described in Section 3 and in Section 4 on a set of typical scenes illustrated in Figure 1.

##### 5.1. Setup

Test scenes are composed of  $M$  media elements. These media elements may be used as several instances in the scene, of which  $N$  are adaptable instances. Depending on designer wishes, a multimedia scene may have  $K$  key scene states corresponding to spatial resolutions where new media elements are introduced. For adaptation techniques with finite granularity,  $P$  additional intermediate states may be created between two key states leading to  $K+P*(K-1)$  states.

It is important to note that the quantitative criteria described in Section 2.2 (*processing*, *memory*, and *bandwidth*) are overheads. They may depend on the number of adaptable media instances, the number of adaptation states and the  $R$  ratio between the weights (in Kbytes) of media data (images, text, and graphics) compared to scene structures. T-DMB interactive scenes can be visualized in Figure 1 and on <http://icme2008.tnt.uni-hannover.de/1835/>. Their characteristics are summarized in Table 1.

Table 1 - Characteristics of test scenes

Sequence	(a)	(b)	(c)	(d)	(e)
$M$	8	13	16	31	39
$N$	8	19	9	19	58
$K$	2	3	3	3	3
$R$	2%	21%	2%	2%	25%

The multimedia scenes have been described using MPEG-4 BIFS and measurements have been performed on Smartphone SPV C500 using GPAC Project [8]. Multimedia scenes with the same visual behavior have been created for each adaptation approach and adaptation algorithms have all been implemented in JavaScript in order to have a fair comparison of processing requirements, even though some approaches could easily be implemented natively (e.g. adaptation scene updates). Finally, it should also be noted that the initialization cost for the JavaScript engine is omitted from these results and that the adaptation processing time corresponds to a complete presentation cycle (including decoding, compositing and rendering).

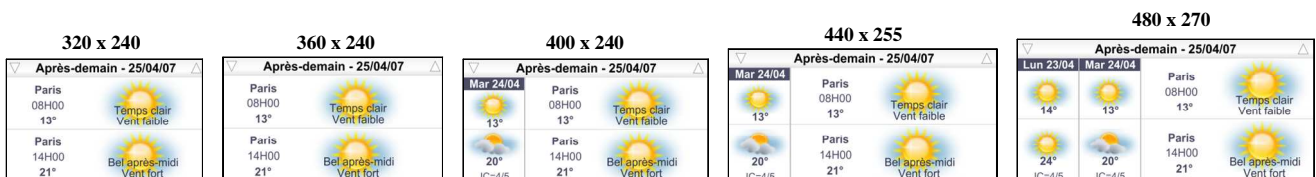


Figure 3 - Possible adaptation for test scene (e) with different resolutions

## 5.2. Results

In terms of adaptation *processing* overhead, Figure 4 depicts a comparison of interpolations, alternatives and proposed scene updates with  $P$  intermediate states on scene (e), illustrated in Figure 3 for three key states ( $K=3$ ). It can be noted that interpolations processing is quite CPU demanding compared to others and that the handling of additional scene structures in the switching approach has a significant impact when  $P$  increases. Figure 4 also shows that the scene updates approach has progressive CPU requirements contrary to the alternatives technique with a similar  $P$  value. Therefore, adaptation using scene updates is a very competitive approach for low-end terminals with a small resolution while it is equivalent to other approaches on high resolution terminals. As explained in Section 5.1, results below are used for a fair comparison and could be optimized.

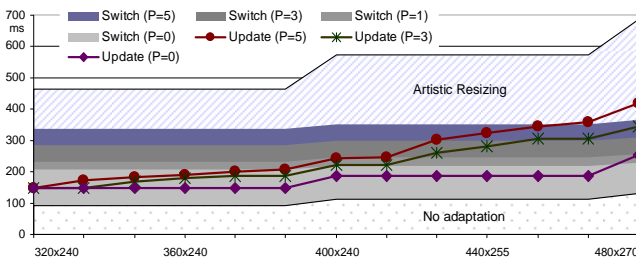


Figure 4 - Adaptation processing overhead on (e)

Concerning *memory* requirements, scene updates do not incur any overhead contrary to other options. Table 2 gives memory overheads for several test scenes. Multimedia scenes with significant scene structures (see Table 1) highlight the limitations of spatial adaptation through alternatives.

Table 2 – Memory adaptation overheads

Sequence	(a)	(b)	(c)	(d)	(e)
Updates	0%	0%	0%	0%	0%
Interpolations	1%	5%	0.9%	2%	<b>12%</b>
Alternatives (P=3)	2%	<b>10%</b>	1.8%	3.7%	<b>21%</b>
Alternatives (P=5)	3%	<b>15%</b>	2.6%	5.3%	<b>30%</b>

The *bandwidth* required to broadcast some adaptable multimedia services is illustrated in Figure 5. The scene is broadcasted every 500 ms (carousel period) and no transport overhead is considered. The benefits of all approaches compared to simulcast are obvious and it can be seen that interpolations and scene updates roughly consume the same bandwidth when few intermediate states are defined ( $P < 3$ ). Furthermore, thanks to the use of our update-based approach, a bandwidth-optimized scene broadcast (noted as Update\* in Figure 5) can even be defined using two different carousels. A base carousel only repeats data for key adaptation states (every 500 ms) while an enhancement carousel provides intermediate adaptation states (every 2 s).

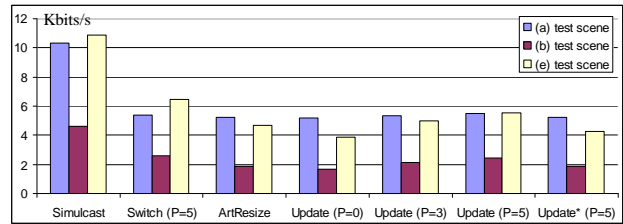


Figure 5 - Bandwidths of adaptable multimedia scenes

## 6. CONCLUSION

This paper has presented how scene updates can be used to achieve spatial adaptation of multimedia services in broadcast environments. Experimental results show that spatial adaptation through scene updates has excellent performances on constrained handsets with small display, low processing power and little memory. Additionally, adaptation scene updates grant content providers with optimized fine-grained spatial adaptation or efficient coarse-grained spatial adaptation depending on acceptable bandwidth costs. In future work, we plan to extend the scope of our experiments and exploit this adaptation technique based on scene updates with other parameters of multimedia scenes temporal or interactive model.

## 7. REFERENCES

- [1] M. Wien et al. "Real-Time System for Adaptive Video Streaming Based on SVC", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1227-1237, September 2007.
- [2] S. Boll, W. Klas, "ZYX – A Multimedia Document Model for Reuse and Adaptation of Multimedia Content", *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 3, June 2001.
- [3] A. Scherp, S. Boll, "MM4U - A framework for creating personalized multimedia content", *Proc. of the Int. Conf. on Distributed Multimedia Systems*, September 2003.
- [4] L. Villard, C. Roisin, N. Layaida, "A XML-Based Multimedia Document Processing Model for Content Adaptation", *Digital Documents and Electronic Publishing*, September 2000.
- [5] G.J. Badros et al. "A Constraint Extension to Scalable Vector Graphics", *Proc. of the 10<sup>th</sup> International World Wide Web Conference*, Hong Kong, 2001.
- [6] S. Yang, J. Zhang, R. Chen, and N. Shao, "A Unit of Information-Based Content Adaptation Method for Improving Web Content Accessibility in the Mobile Internet", *ETRI Journal*, vol. 29, no. 6, pp. 794-807, December 2007.
- [7] M. Kimiaei-Asadi, J.-C. Dufourd, "Context-aware Semantic Adaptation of Multimedia Presentations", *Proc. of ICME*, Amsterdam, Holland, July 2005.
- [8] P. Dragicevic, S. Chatty, D. Thevenin, J.-L. Vinot, "Artistic Resizing: A Technique for Rich Scale-Sensitive Vector Graphics", *Proc. of the 18th annual ACM symposium on User interface software and technology*, pp. 201-210, 2005.
- [9] J. Le Feuvre, C. Concolato, J.-C. Moissinac, "GPAC, Open Source Multimedia Framework", *Proc. of ACM Symposium on Multimedia*, September 2007.