APPLYING MPEG-21 BSDL TO THE JVT H.264/AVC SPECIFICATION IN MPEG-21 SESSION MOBILITY SCENARIOS

Wesley De Neve, Frederik De Keukelaere, Koen De Wolf, and Rik Van de Walle

Multimedia Lab, Department of Electronics and Information Systems, Ghent University, Belgium.
Interuniversity MicroElectronics Center (IMEC), Leuven, Belgium.

ABSTRACT

Video coding is present in a lot of multimedia applications such as video conferencing, digital storage media, television broadcasting, and internet streaming. However, offering universal access to video content is far from trivial when taking into account the diversity of current networks and terminals. In this paper, we will discuss how MPEG-21-based technology can be combined with a recently standardized video compression scheme, called JVT H.264/AVC. To be more specific, we demonstrate how MPEG-21 tools can be used to adapt an H.264/AVC bitstream dynamically to the requirements of a new terminal when transferring a multimedia session between devices. This can be seen as a first step in order to create a framework that allows optimal usage of video data in a network of heterogeneous terminals.

1. INTRODUCTION

In response to the growing need for higher compression of moving pictures, the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) formed a Joint Video Team (JVT) in December 2001 for the development of a new technical specification for digital video coding [1]. Their combined effort was rewarded in the summer of 2003 by the acceptance of a new recommendation by the ITU-T (ITU-T Rec. H.264) and a new international standard by the ISO/IEC (ISO/IEC 14496-10, also known as MPEG-4 AVC or MPEG-4 Advanced Video Coding). The main objectives behind JVT H.264/AVC are an enhanced compression efficiency (provided by the Video Compression Layer or VCL), an improved network adaptation (provided by the Network Abstraction Layer or NAL) and a simple syntax specification.

The specification in question can be seen as the standardized answer to proprietary initiatives and comes in three flavors: the Baseline Profile targeting video conferencing and mobile applications, the Main Profile aiming at broadcasting and entertainment video (such as DVD), and the Extended Profile focusing on streaming. Recent experiments [2] have demonstrated that H.264/AVC's Main Profile offers approximately 40% bit-rate savings relative to the MPEG-4 Visual Advanced Simple Profile and up to 60% bit-rate savings relative to the H.262/MPEG-2 Visual Main Profile, and this for the same quality.

2. PROBLEM DESCRIPTION

Thanks to a reduction in the cost of processing power and memory, and the fast spread of the Internet, an increasing number of people are using network-based multimedia services. Due to the heterogeneity of modern networks and terminals, current multimedia technology has to deal with some major challenges. One of those challenges is the adaptation of video resources when transferring multimedia sessions between those heterogeneous devices, as shown in Figure 1. A new device often implies a new set of constraints, such as a lower available bandwidth, a higher available processing power etc. It is obvious that multimedia formats that are only able to present content with a fixed quality and resolution, are cumbersome nowadays. A delivery system that is based on such formats only reaches a small set of terminals in an adequate manner. Other terminals will receive nothing or will acquire a presentation that is not optimal for their capabilities due to the fact that the content provider is relying on a (lowest) common denominator in order to reach a target audience as large as possible. As such, the use of scalable coding (being able to derive a suitable bitstream by selecting subsets from a parent bitstream) is a must when trying to solve these problems.

In order to make use of the full potential of scalable bitstreams, the definition of a scalable bitstream syntax is insufficient. It is also necessary to develop and use a complete infrastructure that is able to provide an appropriate adaptation and delivery of scalable data such that a diverse public can enjoy an immersive multimedia experience. At this moment, the usage of scalable media often means that content creators have to author several media streams for different types of connections and terminals (a technique better known as simulstore and simulcast). A complete infrastructure, called MPEG-21 [3], that supports the (trans)coding and delivery of

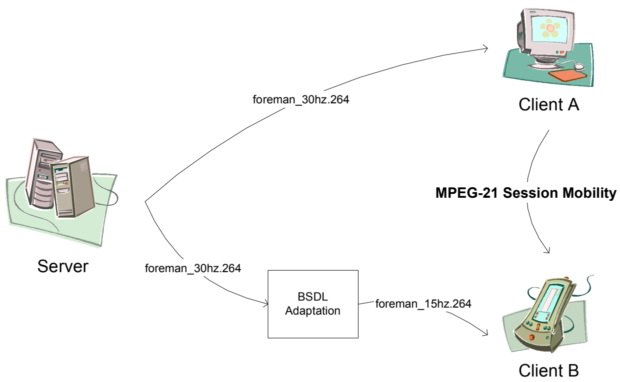

Figure 1: Simplified use case scenario

scalable bitstreams is currently under development and is discussed in the next section.

3. THE MPEG-21 MULTIMEDIA FRAMEWORK

The problem definition as stated above is in line with the vision of the MPEG-21 Multimedia Framework, which is to define a multimedia ecosystem that enables transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. Different MPEG-21-based technologies are currently under development that will allow the creation of such applications. The founders of this framework follow a different approach than was the case for previous MPEG video specifications in the sense that the infrastructure for exploiting scalability is now defined during the first step of the development phase, while a scalable video coding scheme will be fixed during a further stage. The latter procedure will most probably increase the success of a standard for real scalable video coding.

Within MPEG-21 the concept of a "Digital Item" is the key to the whole framework; every transaction, every message, every form of communication is performed by making use of a Digital Item. The latter are defined as "structured digital objects, including a standard representation, identification and metadata". In ISO/IEC 21000-2, the concept of a Digital Item is fully explored and precisely defined. Another part of MPEG-21 which is relevant for this paper, is ISO/IEC 21000-7 [4], better known as Digital Item Adaptation (DIA). Within this part of the MPEG-21 standard, a language based on W3C XML Schema was developed in order to describe the high-level structure of a scalable bitstream, the latter typically comprising a structured sequence of binary symbols. The language in question is known as the Bitstream Syntax Description Language (BSDL) and its strength lies in the fact that it shifts the focus of the adaptation process (required for the delivery of content that is suitable for a certain configuration) from an

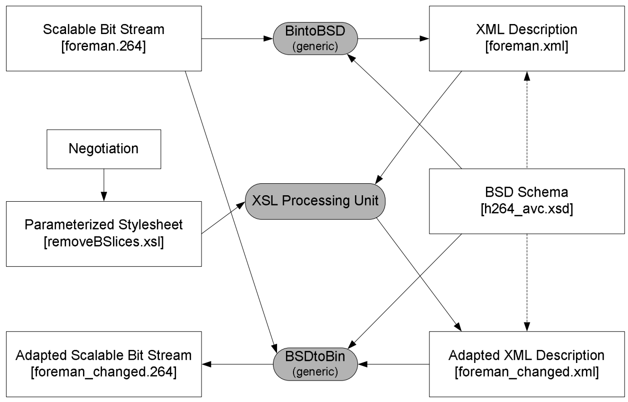

Figure 2: H.264/AVC bitstream adaptation by making use of BSDL

adaptation of the multimedia bitstream to the XML-based description of that particular bitstream, hereby making it possible to create a universal adaptation engine [5]. As shown in Figure 2, the adaptation engine only needs to be instantiated with a BSDL scheme describing a specific bitstream syntax, and a particular style sheet implementing the requested adaptation.

In the Digital Item Adaptation specification there is also a part that standardizes the information that is needed to transfer a multimedia session from one device to another device, commonly called MPEG-21 session mobility. This typically involves information about the state of the Digital Item and information about the multimedia resources that are being consumed, such as the position in the media stream, the state of consumption (paused or not), etc.

4. DYNAMIC ADAPTION OF H.264/AVC BITSTREAMS IN A USE CASE SCENARIO

Figure 1 gives an overview of the simplified use case scenario we will use throughout the rest of this paper. This scenario starts when an MPEG-21 Digital Item Declaration (DID) is consumed by client A. The consumption of the DID results in the delivery of an H.264/AVC Extended Profile bitstream to the client by making use of real-time streaming. At a certain point in time, the session on client A needs to be continued on client B. To realize this, a session mobility Digital Item is constructed and transferred to client B. Client B knows that it only supports the H.264/AVC Extended Profile without B slices (for instance, due to processing constraints) and requests such kind of bitstream from the server. The content provider only has the full-featured Extended Profile version of the bitstream at its disposal and needs to downsample the bitstream to one without B

```xml
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS">
    <Item id="item_01">
        <Choice choice_id="frame_rate">
            <Selection select_id="15Hz"/>
            <Selection select_id="30Hz"/>
        </Choice>
        <Component>
            <Condition require="15Hz"/>
            <Resource mimeType="video/x-mpeg-avc"
                      ref="rtsp://server/foreman_15hz.264"/>
        </Component>
        <Component>
            <Condition require="30Hz"/>
            <Resource mimeType="video/x-mpeg-avc"
                      ref="rtsp://server/foreman_30hz.264"/>
        </Component>
    </Item>
</DIDL>
```

Figure 3: Declaration of a Digital Item

slices. Afterwards the transcoded bitstream is sent to client B. In the following sections, we describe how MPEG-21 tools can be used to realize this use case scenario.

4.1. Digital Item Declaration

To create a Digital Item that can be consumed by terminals with different capabilities, several requirements have to be met. First, the Digital Item must be comprehensible for both terminals. At the same time, it must be possible to include resources that can be consumed by both terminals. Using the Digital Item Declaration Language, Digital Items that fulfill these requirements can be constructed. An example of a DID that can be deployed in the presented use case scenario is shown in Figure 3.

4.2. BSDL Description of H.264/AVC Bitstreams

Universal multimedia access, as defined in the context of video, has two components: error resilience and scalability. Regarding error resilience, H.264/AVC supports tools such as random access, recovery point Supplemental Enhancement Information (SEI) messages, slices (a collection of macroblocks), Flexible Macroblock Ordering (FMO), and slice data partitioning. FMO allows the transmission of macroblocks in an order other than raster scan (for instance, by applying interleaving) while slice data partitioning makes it possible to separate critical components (such as motion vectors) from the less critical ones (such as the prediction error). Note that each slice header is strongly related to the resynchronization marker concept as found in MPEG-4 Visual [6], since its syntax – among other things – contains information about the position of the first macroblock that is being conveyed in its corresponding slice data.

Although it was considered desirable in the terms of reference, true bitstream scalability (quality, temporal, spatial, or fine grained) was not achieved or explored during the development of the H.264/AVC standard (which was in fact expected by the developers). However, the H.264/AVC specification does allow seamless switching between streams of different rates (by means of switching pictures), which provides more or less the same functionality for many services and applications. In addition to that, slice data partitioning can also be seen as a trivial implementation of quality scalability and one of the purposes of our research activities is to exploit this type of scalability by making use of BSDL.

So far, we have already developed a BSD schema for bitstreams compliant to the Byte Stream NAL unit syntax. This schema supports all tools as defined in H.264/AVC's Main Profile. By relying on a non-normative feature of BSDL, we were able to extend the language in question by an additional data type, i.e. the exp-Golomb data type. The exponential Golomb code is a variable length code with a regular construction, and this code is frequently used in the H.264/AVC specification for the representation of header information (including the syntax element `slice_type`, as depicted in Figure 4. As such, our schema allows the description of a bitstream up to the level of slice headers. By making use of a stylesheet that examines the value of the `slice_type` symbol, we are able to drop bidirectional predicted slices (B slices), thus implementing a trivial form of temporal scalability. However, since B slices can be used in H.264/AVC for the prediction of other slices, they cannot be dropped randomly. The latter implies that some care has to be taken into account during the encoding process when a content provider wants to exploit temporal scalability in the H.264/AVC specification. Note that it is most probably even possible to transcode an H.264/AVC bitstream, compliant to the Main Profile, to a bitstream that satisfies the Baseline Profile constraints (among other things not allowing B slices) when the above approach is applied. The latter illustrates another application area for BSDL. However, in this example one will also have to take into account some important side effects, such as signaling the appropriate profile and level information in a Sequence Parameter Set. This latter change could also be applied by the stylesheet used for dropping B slices.

4.3. MPEG-21 Session Mobility

A third technology, called session mobility, of the MPEG-21 standard that is used in our scenario is also part of the Digital Item Adaptation specification. This specification allows a standardized description of session information.

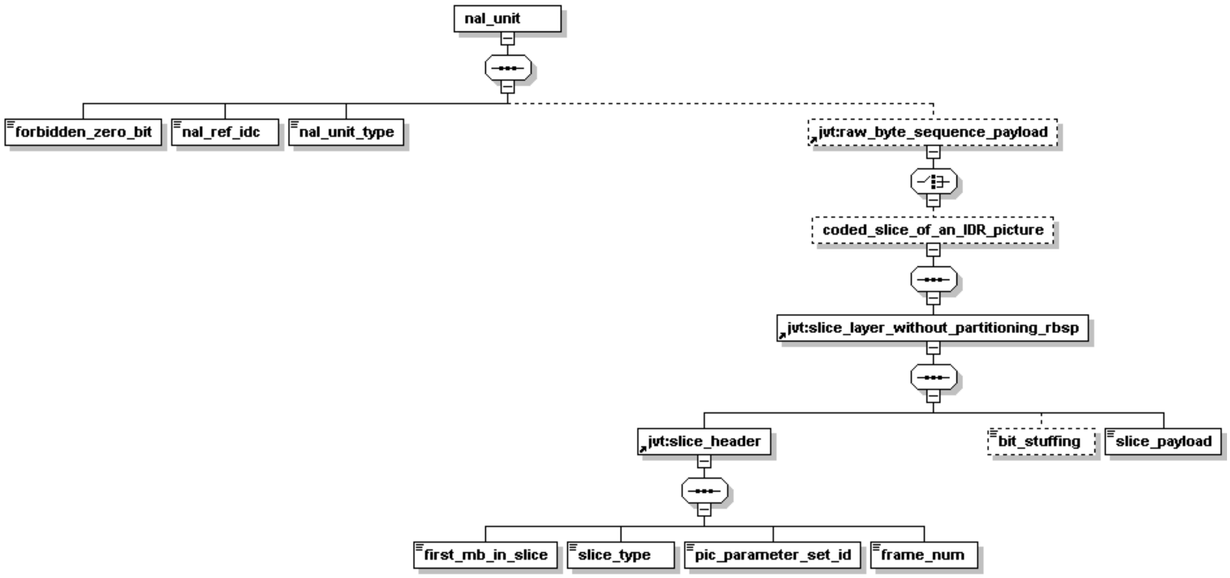

Figure 4: Snapshot of the H.264/AVC syntax (irrelevant syntax elements are omitted)

This session information can be transferred to other devices where it is used to reconstruct sessions on those new devices. During the reconstruction of the multimedia session, BSDL, or even any resource adaptation tool, can be applied to adapt the resources to the requirements of the new terminal.

5. CONCLUSION

In this paper, we have presented how to exploit temporal scalability in the JVT H.264/AVC specification, by making use of MPEG-21 BSDL. In order to achieve this, we have constructed a BSDL schema for the JVT H.264/AVC specification up to the level of slice headers. This BSDL approach has the advantage that existing XML tools can be used to adapt bitstreams to new bitstreams suited for consumption by devices with a wide range of terminal and network characteristics. Using this approach, we have demonstrated how MPEG-21 Session Mobility can merit from resource adaptation in general or BSDL in specific.

However, further research can be done to know whether BSDL is a feasible tool for exploiting scalability in the JVT specification. Several topics still need to be explored. For instance, the exact conditions under which B slices may be dropped and whether BSDL can cope with emulation prevention bytes (which are escape codes in H264/AVC bitstreams).

6. ACKNOWLEDGMENTS

The research activities that have been described in this paper were funded by Ghent University, the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC), and the European Union.

7. REFERENCES

[1] T. Wiegand, G. Sullivan, and A. Luthra, “Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)”, ISO/IEC JTC1/SC29/WG11 and ITU-T VCEG, Geneva, 2003.

[2] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, “Rate-constrained coder control and comparison of video coding standards,"' IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp. 688-703, July 2003.

[3] I. Burnett, R. Van de Walle, K. Hill, J. Bormans, and F. Pereira “MPEG-21: Goals and Achievements”, *IEEE Multimedia,* IEEE Computer Society, pp. 60-70, 2003

[4] Moving Picture Experts Group, “Text of ISO/IEC 21000-7 FCD – Part 7: Digital Item Adaptation,” ISO/IEC JTC1/SC29/WG11 N5845, July 2003

[5] Myriam Amielh, Sylvain Devillers, “Bitstream Syntax Description Language: Application of XML-Schema to Multimedia Content Adaptation,” WWW2002, May 2002

[6] Moving Picture Experts Group, “MPEG-4 Video Verification Model 18.0 (VM-18),” ISO/IEC JTC1/SC29/WG11 N3908, January 2001