Applying MPEG-21 BSDL to the JVT H.264/AVC Specification for Digital Video Coding

Wesley De Neve, Frederic Van Quickenborne, Ingrid Moerman, Piet Demeester, Rik Van de Walle

Abstract— **Video coding is used under the hood of 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. This can be seen as a first step in order to create a framework that allows optimal usage of video data under constraining circumstances.**

Keywords— **Multimedia, MPEG, Video, H.264, AVC, Scalability, Compression, BSDL.**

I. 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 by the acceptance of 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 focussing 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.

II. 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 is making use of network-based multimedia services. Due to the heterogeneity of modern networks and terminals, current multimedia technology has to deal with some major challenges. It is obvious that multimedia formats, only able to present content with a fixed quality and resolution, are cumbersome nowadays. A delivery system that is based on such formats will only reach a small number of interested users in an adequate manner. Other users will receive nothing or will acquire a presentation that is not conformant with the capabilities of their network connections and terminals. As such, the use of scalable coding (being able to derive useful video from subsets of a 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 format is insufficient. It is also necessary to develop and to make use of 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). Complete infrastructures that support the (trans)coding and delivery of scalable bitstreams do not exist yet.

III. Description of the research

A. Description and Manipulation of Scalable Bitstreams

The problem definition as stated above is in line with the vision of the MPEG-21 Multimedia Framework, which is to define a multimedia framework that enables transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. As such, among other technologies, a language based on W3C XML Schema was developed within MPEG-21 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 Definition Language (BSDL) and its strength lies in the fact that it shifts the focus of the adaptation proces (required for the delivery of content that is suitable for a certain configuration) from an adaptation of the multimedia bitstream to the XML-based description of that particular bitstream, hereby making it possible to create a universal adaptation engine. As shown in Fig. 1 the adaptation engine only needs to be instantiated with a BSDL scheme describing a specific bitstream syntax, and a particular stylesheet implementing the re-

W. De Neve and R. Van de Walle are with Multimedia Lab, ELIS Department, Ghent University-IMEC, Ghent, Belgium. E-mail: {wesley.deneve, rik.vandewalle}@ugent.be.

F. Van Quickenborne, I. Moerman and P. Demeester are with IBCN, INTEC Department, Ghent University-IMEC, Ghent, Belgium. E-mail: {frederic.vanquickenborne, ingrid.moerman, piet.demeester}@ugent.be

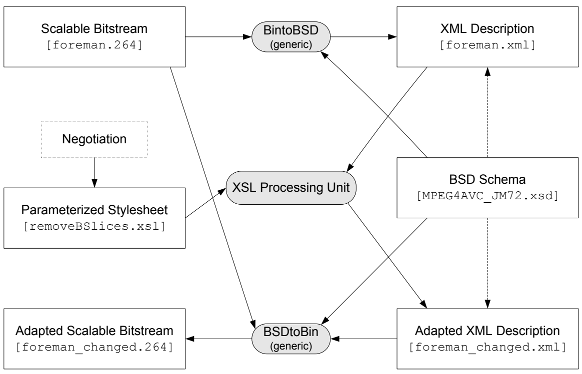

Fig. 1. H.264/AVC bitstream adaptation with BSDL.

```
bitstream
  byte_stream
    byte_stream_nal_unit
      zero_byte
      start_code_prefix_one_byte
      nal_unit
        forbidden_bit
        nal_ref_idc
        nal_unit_type
        raw_byte_sequence_payload
          ...
          coded_slice_of_a_non_IDR_picture
            slice_layer_without_partitioning_rbsp
              slice_header
                first_mb_in_slice
                slice_type
                ...
              slice_id
              bit_stuffing
              slice_payload
          ...
```

Fig. 2. Snapshot of the Byte Stream NAL unit syntax.

quested adaptation. One of the other advantages of describing the syntax of a bitstream using XML is that there are already quite some solutions available to perform XML-to-XML editing operations.

B. Universal Multimedia Access and JVT H.264/AVC

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).

Although it was considered as 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 specification (which was in fact expected by the developers). However, the H.264/AVC codec does allow seamless switching between streams of different rates (by the introduction 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[1] 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 datatype, i.e. the exp-Golomb datatype. The exponential Golomb code is in fact a variable length code with a regular construction, and this coded is frequently used in the H.264/AVC specification for the representation of header information (including the syntax element `slice_type`, as depicted in Fig. 2). 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 bidirectionally 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.

[1] Nowadays, quality scalability is often replaced by multi-step post-filtering at the decoder.

IV. Conclusion

In this paper we have presented some ideas we have on how to exploit temporal and quality scalability in the JVT H.264/AVC specification, and this by making use of MPEG-21 BSDL. One of the main advantages of our approach is the fact that it can be built on top of the existing syntax and tools. However, further research will be necessary in order to know the exact conditions under which B slices may be dropped and in order to know whether BSDL is a feasible tool for achieving quality scalability. We will also pay attention to future proposals, such as [3] in which an SNR-scalable extension for H.264/AVC is discussed, and how these suggestions can be linked to BSDL.

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.

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] H. Schwarz, D. Marpe, and T. Wiegand, *SNR-scalable Extension of H.264/AVC*, ISO/IEC JTC1/SC29/WG11 and ITU-T VCEG, San Diego, 2003.