FORMAT-INDEPENDENT SCALABLE BIT-STREAM ADAPTATION USING MPEG-21 DIA

Debargha Mukherjee, Geraldine Kuo, Shih-ta Hsiang, Sam Liu, Amir Said

E-mail: debargha@hpl.hp.com, Geraldine.Kuo@hp.com, hsiang@ieee.org, lius@hpl.hp.com, said@hpl.hp.com
Hewlett Packard Laboratories, Palo Alto, CA 94304.

ABSTRACT

Part 7 of MPEG-21 entitled Digital Item Adaptation (DIA), is an emerging metadata standard defining protocols and descriptions enabling content adaptation for a wide variety of networks and terminals, with emphasis on format-independent mechanisms. The DIA descriptions provide a standardized interface not only to a variety of format-specific adaptation engines, but also to a fully format-independent adaptation engine for scalable bit-streams. A format-independent engine contains a decision-taking module operating in a semantics- independent manner, cascaded with a bit-stream adaptation module that uses an XML transformation to model the bit-stream adaptation process using parameters derived from decisions made. In this paper, we describe the DIA descriptions that enable such fully format-independent bit-stream adaptation. Universal adaptation engines substantially reduce the adoption costs because the same infrastructure can be used for different types of scalable media, including proprietary and encrypted.

1. INTRODUCTION

Network and terminal capabilities continue to grow, as does the disparity of networks and terminals that need to coexist. There is also fast growth of the richness of delivered content, presenting a formidable obstacle to universal media access. Under these circumstances a rigid content format is clearly inappropriate because it caters only to a small subset of potential recipients. Scalable formats, on the other hand, enable efficient and secure adaptation [1][2][3][4] by allowing downscaling of content by simply deleting segments from the bit-stream, along with minor bit-stream editing. Currently, one fully scalable and efficient image-encoding standard available is JPEG2000 [5]. A standard on fully scalable video (MPEG-21 Part 13) is under development. Note however that all forms of MPEG video are already temporally scalable, and in addition support various other modes of scalability, for instance MPEG-4 FGS [6]. However, despite the distinct advantages of efficient and secure adaptation that scalability promises, its industry adoption has been slow because different protocols and adaptation engines are needed for different types of content. Since any infrastructure is expensive to deploy, the benefits of scalability have been overshadowed by the cost of adoption. The alternative is to develop *format-independent* infrastructures, which need to be deployed only once to support content adaptation and delivery of a broad range of media types, present and future. They can play a key role in successful adoption of scalable bit-streams, unlocking their true potential. They also enable instant delivery of new proprietary scalable formats without additional investment in infrastructure.

The emerging MPEG-21 standard [7] defines a framework to enable transparent use of multimedia resources across a wide range of networks. Part 7 of MPEG-21, entitled Digital Item Adaptation (DIA) [8][9], deals with descriptions and protocols enabling content adaptation to customize for a wide variety of networks and terminals, with emphasis on format independent mechanisms. DIA was approved for FDIS in December 2003. A proposal for DIA, entitled Structured Scalable Metaformats (SSM) [1][2][10], developed a comprehensive end-to-end framework for scalable bit-stream adaptation. SSM was originally proposed [1] as a meta-format for scalable bit-streams, and consisted of standardized headers that a universal adaptation engine can interpret and take action on and enforced various restrictions on the bit-stream. Later, in order to match the goals of MPEG-21 DIA, the header information was recast as an XML descriptor accompanying a bit-stream, with various enhancements to obviate the need for explicit restrictions on the format [10]. During the standardization process, many SSM concepts were adopted into the DIA standard either normatively or informatively, while some others are queued for further study. In this paper, we describe the concepts behind many components in DIA, with focus on those that enable fully format-independent adaptation of scalable bit-streams, as envisioned in SSM.

2. FORMAT-INDEPENDENT ADAPTATION

The high-level adaptation architecture of an adaptation engine envisioned by SSM and DIA is shown in Figure 1. Figure 1(a) depicts the external model with the following interface to an engine:

Inputs:

- An input bit-stream,
- An input bit-stream description, consisting of various kinds of metadata related to the content. This input bit-stream description is typically generated by the content creator and travels with the bit-stream.
- An outbound constraints description, specifying the network and terminal constraints that must be satisfied by the outbound adapted bit-stream. This description is generated either by the recipient, or a network sensor, or a license provider.

Outputs:

- The adapted bit-stream, and
- The adapted bit-stream description (optional) for use in a subsequent stage of adaptation.

In a format-independent scalable bit-stream adaptation engine a generic processing model for the adaptation engine can be derived for any scalable bit-stream. Descriptions change with the bit-stream, but the generic processing of the descriptions and the bit-stream handles the adaptation process, as required to provide the right adapted bit-stream, irrespective of the format. Figure 1(b) shows a more detailed view of such an adaptation engine. There is an Adaptation Decision Taking Engine (ADTE) module that processes the outbound constraints and a part of the input bit-stream description to make adaptation decisions. The decisions are next input to a Bit-stream and Description Adaptation Engine (BDAE) to perform the bit-stream adaptation, using other information from the bit-stream description as required.

In the MPEG-21 framework [7], a bit-stream and all its associated descriptions are packaged as what is called a Digital Item. The relevant descriptions for adaptation, called tools, are standardized

0-7803-8554-3/04/$20.00 ©2004 IEEE.

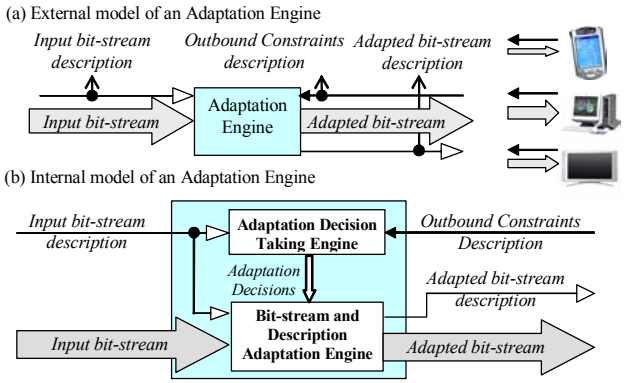

Figure 1. Adaptation Engine model.

in Part 7, entitled Digital Item Adaptation (DIA) [8][9]. Thus, if the adaptation architecture depicted in Figure 1 were implemented using MPEG-21, the inputs and outputs would be packaged as Digital Items (DI). Further, the input bit-stream description and the outbound constraints descriptions of Figure 1 are represented using several DIA tools, specialized by functionality.

In particular, for the outbound constraints description input to the adaptation engine, there are two ways to provide that using DIA tools: the Usage Environment Description (UED), and the Universal Constraints Description (UCD). The UED is a major part of the standard providing standardized descriptions of network and terminal capabilities that can be used by a variety of adaptation engines. Examples include network characteristics, the display, audio, or video capabilities of a terminal, etc. While these descriptions are easy to process because they have a definite structure, they are not directly useful for format-independent adaptation since the processing engine in this case has to specifically understand the description semantics, and make reasonable assumptions about how that translates to constraints to be applied for adaptation. A truly format independent engine should not require any assumption on the type of content the bit-stream represents. The UCD provides an alternative that represents constraints on content characteristics directly. The UCD can in turn reference values from the UED, but the processing driven by the UCD can be semantics independent.

There are several DIA tools that must be used for input bit-stream description. The information used solely in the ADTE module—to make appropriate adaptation decisions based on outbound constraints—is packaged in a DIA tool called the AdaptationQoS (AQoS). There are tools called Bit-stream Syntax Description (BSD) that provide a high-level description of the bit-stream syntax at the granularity needed for adaptation. The BSDLink tool maps decisions to parameters used in a stylesheet (using any transformation language), to be applied to the BSD to specify the bit-stream adaptation mechanism. DIA has chosen not to standardize any specific transformation for this purpose, but there is a proposal for this based on scalable bit-stream models.

In the next two sections, we present the concepts behind the ADTE and the BDAE modules, with focus on the DIA tools that are processed in these modules.

3. DECISION-TAKING

The task of the ADTE is to take appropriate decisions for adaptation of an input bit-stream subject to constraints for the outbound adapted content. To ensure format-independence, this operation must be semantics-independent. If a universal adaptation engine is to take decisions based on certain constraints, without knowledge or assumption whatsoever about either the content or what the constraints represent, the semantics must be abstracted, and the decision-taking problem must be cast in mathematical language.

In SSM and MPEG-21 DIA decision-taking is defined as a constrained optimization problem involving *variables* that represent content or usage environment characteristics. The DIA tools, AdaptationQoS and UCD, used in combination, create a mechanism for semantics-independent decision-taking for bit-streams covering a wide range of useful adaptation scenarios. The UCD can still reference values from the UED, but the processing does not require understanding of their semantics. Note that while it is possible to use the AdaptationQoS and UCD tools independently in a format-specific engine, in this paper we focus on the fully semantics-independent decision-taking mechanism enabled by the standard.

The framework enables decision-taking not only on the bit-stream as a whole, but differentiated with respect to logical segments corresponding to partitions such as GOPs, ROIs, Tiles, Frames, etc., referred to as *adaptation unit* in DIA. The variables for the optimization problems are defined on a per adaptation unit basis, and are termed IOPins in the AdaptationQoS description. While some of the variables may correspond purely to usage environment and preference inputs, some others may correspond to resource characteristics with respect to available adaptation options differentiated by adaptation units, and yet others may combine resource characteristics with usage environment/preference inputs. The set of variables for the nth adaptation unit is denoted by vector $\mathbf{I}[n] = \{i_0[n], i_1[n], .., i_{M-1}[n]\}$, n = 0, 1, 2, …, where M is the total of variables, and n is the adaptation unit index.

The AdaptationQoS tool declares and defines vectors $\mathbf{I}[n]$, provides information on the possible values they take, and also conveys the interdependencies between them, the UCD tool specifies the actual optimization problem for each adaptation unit. In particular, the UCD conveys for each adaptation unit n, numeric expressions $O_{n,j}(\mathbf{I}[n], \mathbf{H}[n])$, j=0,1,…,J_n–1 called optimization constraints, along with several Boolean expressions $L_{n,k}(\mathbf{I}[n], \mathbf{H}[n])$, k=0,1,…,K_n–1, called limit constraints, which are used together to specify the following optimization problem involving $\mathbf{I}[n]$:

Maximize $\{O_{n,j}(\mathbf{I}[n], \mathbf{H}[n])\}$, $j = 0, 1, ..., J_n-1$

subject to: $L_{n,k}(\mathbf{I}[n], \mathbf{H}[n])$ = `true`, $k = 0, 1, ..., K_n-1$

The number of optimization constraints (J_n) is arbitrary. If J_n=0, any solution in the feasible region is acceptable. The case J_n=1 is the most typical and defines a single-criterion optimization problem with usually a unique solution. The case J_n>1 defines a multi-criteria optimization problem [11], where any solution in the Pareto optimal set region is sought.

Let I*[n] represent a solution for the nth adaptation unit. The vector H[n] in the expressions of $O_{n,j}$ and $L_{n,k}$ represent the history of past decisions for adaptation units 0,1,…, n-1. In other words, H[n] = {I*[0], I*[1],…,I*[n–1]}. A decision module based on the AdaptationQoS and the UCD makes decisions for the vectors I[n] sequentially in order of n = 0,1,2,…

In most cases involving scalable bit-streams, the AdaptationQoS provides information on a discrete set of possible adaptation choices, and how other variable values change based on each choice. In such cases, the optimization problem is readily solved by an exhaustive search over the space of all possible adaptations, but the DIA descriptors do not preclude the use of more sophisticated

optimizers for the same purpose.

Within the ADTE the semantics of the IOPins are immaterial because they are simply regarded as mathematical variables in an optimization problem, but they are very important at the receiver end or other places where the UCD originates. That is because the UCD creator in many scenarios would not be expected to know the identifier of the IOPin (variable) defined in the provider side AdaptationQoS description, corresponding to a given semantics. In order to enable linking of the UCD to the right IOPins in AdaptationQoS, DIA creates a number of classification schemes to standardize terms having pre-defined semantics for representing media characteristics, usage environment characteristics, and segment decompositions. The AdaptationQoS tool associates the IOPins it defines with these terms, while the UCD creator uses the semantics terms to specify the optimization problem, rather than use the identifier of the IOPins directly. The ADTE simply needs to do a match of the classification scheme terms used in AdaptationQoS and UCD to know how the constraints specified in UCD map to constraints for IOPin variables. The use of classification scheme terms for the purpose of semantic linking is conceptually in line with Reserved variables envisioned in the SSM framework.

4. BITSTREAM AND DESCRIPTION ADAPTATION

Once optimal decision solutions **I***[*n*] have been obtained by the ADTE, they need to be used for the actual bit-stream adaptation process. In DIA, format-independent bit-stream adaptation is modeled by an XML transformation applied to the Bit-stream Syntax Description (BSD). DIA tools BSD and its variant generic BSD (gBSD), describe the high level structure of a bit-stream in XML form. Using transformation operations specified in an XML transformation stylesheet, to be applied to the (g)BSD, the BDAE can transform the bit-stream and the corresponding description using editing-style operations such as data truncation and simple modifications that are sufficient for scalable bit-stream adaptation. The actual language to be used for the transformation stylesheet is non-normative in DIA.

The decisions made by the ADTE are linked to the BDAE by passing parameters to the transformation stylesheet. The BSDLink tool specifies how decisions map to parameters. Since different decisions **I***[*n*] are made for different adaptation units, the stylesheet parameters are vectors in the general case, where the vector index corresponds to successive adaptation units. The adaptation of a bit-stream segment belonging to a given adaptation unit, is governed by parameters for the corresponding adaptation unit. The BSD and the stylesheet needs to be designed jointly, such that it is possible to know which segments belong to which adaptation units.

The generic BDAE architecture is shown in Figure 2. Here the XML Transformation Processing + Bit-stream Adaptation block may be implemented as a cascade of a generic XML transformation module (such as XSLT processor if a XSLT stylesheet is used), followed by a bit-stream generator. On the other hand, in the interest of efficiency the transformation and the bit-stream generation processes may and should be combined.

DIA references W3C XSLT or SourceForge's STX as possible general-purpose languages to use for the transformation stylesheet, but proprietary transformations may be used as well. A model based XML transformation language entitled the BSD Transformation Instructions (BSDTrI), was proposed in DIA based on the scalable bit-stream modeling concepts in SSM. SSM establishes a

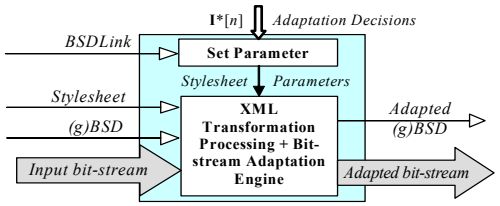

Figure 2. Bit-stream and Description Adaptation Engine

universal model for all scalable bit-streams, referred to as the SSM bit-stream model, which not only provides the set of possible adaptation choices unambiguously, but also how to adapt a scalable bit-stream by segment deletions. This model was later reused by Lerouge et al [11]. Other micro-models handle automatic, efficient update of certain fields in the bit-stream besides segment deletions. The BSDTrI is an XML transformation language created based on these models in SSM. The advantages of a language specific for scalable bit-streams using higher level bit-stream modeling, over a lower level general purpose XML transformation language such as XSLT, are: efficiency – both in terms of processing time and memory requirement, compactness of the descriptor, ease of implementation on a custom or embedded device because of a limited instruction set, and ease of creation. The core experiments conducted in MPEG so far validate these claims, but further study is required to better understand the proposed language and refine its specification. The current specification of the BSDTrI is in AM 8.0 [9].

5. ENGINE IMPLEMENTATION AND USE CASES

A fully format-independent adaptation engine based on the described DIA tools was created in course of the MPEG-21 DIA standardization process. Figure 3 presents the architecture of this engine and illustrates how DIA tools are used in the ADTE and BDAE modules. The input bit-stream and the associated descriptions are packaged as an input DI originating from the content provider. The outbound constraints input is comprised by the UCD with possible references to UEDs, that originate from the recipient or other network nodes. The output bit-stream and descriptions are again packaged as a DI. Note that the current MPEG-21 specification supports adaptation of the (g)BSD but not the AdaptationQoS. Without this, the output DI cannot be used for a subsequent stage of adaptation. This feature should be pursued in an amendment.

The adaptation engine software and some use cases are available at [12]. In this paper, we present one use case involving MC-EZBC [13] – a fully scalable video codec proposed for the Scalable Video Coding standard MPEG-21 Part 13. Such a bit-stream is organized into multiple sequentially transmitted groups of frames (GOF) that constitute adaptation units, each containing typically 16 or 32 frames. Each GOF is coded into several temporal, spatial and SNR layers. This model is likely to remain the same no matter what fully scalable video format is used.

An ADTE makes decisions on the number of temporal, spatial and SNR layers to include for each successive GOF (adaptation unit), based on current network and terminal constraints. For streamed content, the ADTE is designed to make decisions for successive GOFs (adaptation units) synchronously with the transmission schedule, in order to accommodate dynamically changing network and terminal conditions. In other words, the UCDs and UEDs that actually provide the constraints may change dynamically during a streaming session, causing the ADTE decisions for

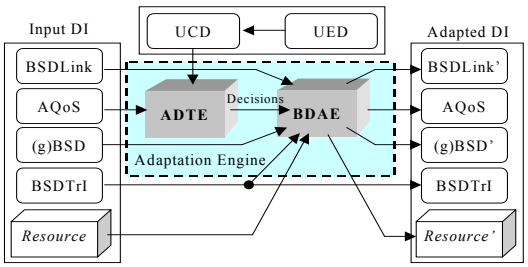

Figure 3. Fully format-independent adaptation software

the currently processed and transmitted GOFs to also change accordingly to accommodate them.

The content provider makes available the AdaptationQoS metadata that defines and declares the following IOPins:

- Adaptation unit IOPin: `GOF`.
- Free IOPins: `NTEMP`, `NSPATIAL`, `NSNR` – indicating number of temporal, spatial, and SNR layers respectively.
- Dependent IOPins: `FRAMERATE` (temporal resolution), `BITRATE` (rate), `PQUAL` (perceptual GOF quality combining frame SNR computed at the highest resolution with framerate), `FRAMEWIDTH`, `FRAMEHEIGHT` – each as a function of free IOPins per GOF.

The following example UCD requests an adaptation:

- For the first adaptation unit (GOF), the UCD requests: Maximize `PQUAL`, subject to: `FRAMEWIDTH` ≤ *display width* provided; `FRAMEHEIGHT` ≤ *display height* provided; `FRAMERATE` ≥ *a minimum desired value*; `BITRATE` ≤ *average transmission rate supported by network.*
- For all subsequent adaptation units, the `FRAMEWIDTH` and `FRAMEHEIGHT` limit constraints are replaced by one, which requires the `FRAMEWIDTH` (or `FRAMEHEIGHT`) to remain the same as that for the previous adaptation unit.

Thus, the ADTE chooses the spatial resolution only for the first GOF, and maintains it the same for all subsequent GOFs. However, the temporal and SNR layers chosen keep changing depending on the video characteristics as provided in the AdaptationQoS and the current network conditions.

The adaptation performance for this use case is demonstrated on 288 frames of the CIF *Foreman* sequence, compressed using the MC-EZBC [13] inter-frame scalable video codec. The compressed bit-stream consists of 18 16-frame GOFs, each with 5 temporal, 6 spatial and 5 SNR layers. The parameters provided in the UCD generate a QCIF resolution adapted video for the first GOF, which is maintained for all subsequent GOFs. We consider two cases: first where the average available transmission rate supported by the network is 700 Kb/s for the entire duration of the transmission, and the second where the constraints are dynamically updated every one-third of the video so that for the same average rate of 700 Kb/s, the available rates for each individual one-third of the video are 700 Kb/s, 350 Kb/s and 1050 Kb/s respectively. Table 1 presents for both cases the actual bandwidth transmitted along with the number of temporal, spatial and SNR layers transmitted for each GOF.

The results presented use an empirical measure of perceptual quality `PQUAL`. In the context of fully scalable video, we believe there needs to be a comprehensive psycho-visual study on the perceptual characteristics of temporal resolution combined with frame SNR, in order to derive a measure of quality or distortion per GOF, and make an educated adaptation decision.

Table 1. Dynamic adaptation to match available bandwidth. T/S/Q represents Temporal/Spatial/SNR(Quality) layers preserved. All BWs are in Kb/s.

GOF	Constant BW			Dynamic BW		
	Av. BW	Actual BW	T/S/Q Layers	Av. BW	Actual BW	T/S/Q Layers
0	700	600	5/5/2	700	600	5/5/2
1	700	671	4/5/3	700	671	4/5/3
2	700	536	5/5/2	700	536	5/5/2
3	700	544	5/5/2	700	544	5/5/2
4	700	542	5/5/2	700	542	5/5/2
5	700	670	5/5/2	700	670	5/5/2
6	700	657	5/5/3	350	332	4/5/2
7	700	679	3/5/4	350	273	5/5/1
8	700	633	5/5/2	350	321	4/5/1
9	700	669	5/5/2	350	317	4/5/1
10	700	579	5/5/2	350	290	4/5/1
11	700	651	5/5/2	350	308	4/5/1
12	700	687	4/5/3	1050	889	5/5/3
13	700	521	5/5/2	1050	870	5/5/3
14	700	607	4/5/3	1050	978	4/5/4
15	700	665	5/5/3	1050	876	4/5/4
16	700	587	5/5/3	1050	827	4/5/4
17	700	553	5/5/3	1050	1010	4/5/4

6. REFERENCES

[1] D. Mukherjee and A. Said, "Structured Scalable Meta-formats (SSM) for Digital Item Adaptation," *Proc. SPIE, Internet Imaging IV*, vol. 5018, pp. 148-67, Jan 2003.

[2] D. Mukherjee, P. Chen, S-T. Hsiang, J. Woods, and A. Said, "Fully Scalable Video Transmission using the SSM Adaptation Framework," *Proc. SPIE, Visual Commun. and Image Proc.*, vol. 5150, July 2003.

[3] S. J. Wee and J. G. Apostolopoulos, "Secure scalable video streaming for wireless networks," *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing*,vol. 4, pp. 2049–2052, May 2001.

[4] S. J. Wee and J. G. Apostolopoulos, "Secure scalable streaming enabling transcoding without decryption," *Proc. IEEE Int. Conference on Image Processing,* vol. 1, pp. 437–440, Oct. 2001.

[5] D.S. Taubman and M.W. Marcellin, "JPEG2000: Image Compression Fundamentals, Standards and Practice," *Kluwer Acad. Pubs, 2002*.

[6] Weiping Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, pp. 301–317, March 2001.

[7] J. Bormans, J. Gelissen, A. Perkis, *MPEG-21: The 21st century multimedia framework*, Signal Processing Magazine, IEEE, Volume 20, Issue 2, pp. 53 – 62, March 2003.

[8] "ISO/IEC 21000-7 FDIS Part 7: Digital Item Adaptation," *ISO/IEC JTC 1/SC 29/WG 11/N6168*, Dec 2003, Hawaii, USA.

[9] "MPEG-21 Digital Item Adaptation AM (v8.0)", *ISO/IEC JTC 1/SC 29/WG 11/N6169*, Dec 2003, Hawaii, USA.

[10] Debargha Mukherjee et. al. "Structured scalable meta-formats version 1.0 for content agnostic digital item adaptation," ISO/IEC JTC1/SC29/WG11 MPEG2002/M9131, Dec 2002.

[11] S. Lerouge, P. Lambert, R. Van de Walle, "Multi-criteria Optimization for Scalable Bitstreams," *Visual Content Processing and Representation*, *8th International Workshop VLBV 2003, Lecture Notes in Computer Science*, vol. 2849, Sept. 2003.

[12] http://www.hpl.hp.com/research/ssm/mpeg21/DIA-FORMAT-INDEP-ADAPTATION-USECASES-HP.v3.zip

[13] S. –T. Hsiang, J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," Signal Processing: Image Communications, vol. 16, pp. 705-24, May 2001.