MPEG-21 Standardization Process: Organization and Rate Distortion Modeling for Network Adaptation

ZORAN BOJKOVIC, BOJAN BAKMAZ
University of Belgrade
SERBIA
z.bojkovic@yahoo.com

Abstract: An overview of MPEG-21 standard together with the application of the rate distortion theory in modeling for network adaptation is provided. The first section introduces parts of this standard as well as motivation and objective for all of them. Next section deals with Digital Item Adaptation (DIA) as a part of the MPEG-21 Multimedia Framework. DIA specifies metadata for assisting the adaptation of digital items according to the constraints on the storage transmission and consumption. In that way various types of quality of service management are enabled. Finally, the third section describes DIA use in multimedia applications and reports on some of the activities in MPEG-21 on extending DIA for the use in the rate distortion modeling for network adaptation.

Key-Words: Digital Item Adaptation, MPEG-21 standard, rate distortion, QoS.

1 Introduction
Standards are technical documents that define precisely the conformance required of users if interoperability is to be achieved. MPEG-21 can be traced back to mid 1999 as an activity within the Moving Picture Experts Group (MPEG) community. Since then, many people worldwide have worked to create a coherent and valuable framework for multimedia delivery and consumption. This has required trust that the original vision was correct. The novelty and different directions of the MPEG-21 work have lead to new and valuable tests for long-standing MPEG process.

MPEG has played a key role in developing the standards behind the explosion of multimedia-enabled consumer devices. The MPEG-1 and MPEG-2 standards are the core technologies behind digital TV, advanced audio coding and DVDs. The MPEG-4 standard has seen success in its use in IP video content, while MPEG-4, part 10 (H.264/AVC) standard is making inroads into the mobile content and broadcasting. MPEG-7 is an extensive standard for the description of multimedia content using eXtensible Markup Language (XML) metadata. Finally, in MPEG-21, the aim is to create a standard that would link together the media coding and metadata standards with access technologies, rights and protection mechanisms, adaptation technology and standardized reporting in order to produce a complete multimedia framework. Thus, MPEG-21 is a major step forward in multimedia standards. It collects together the technologies needed to create an interoperable infrastructure for protected digital media consumption and delivery. MPEG-21 Multimedia Framework has been started with the goal to enable transparent and augmented use of multimedia resources across a wide range of networks and devices. A control concept is that the user is the most important.

This work provides an overview of MPEG-21 standard and contains three sections. The first section introduces the parts of this standard together with motivation and objective for all of them. Next section deals with Digital Item Adaptation (DIA) as part of the MPEG-21 Multimedia Framework. DIA specifies metadata for assisting the adaptation of digital items according to the constraints on the storage transmission and consumption, thereby enabling various types of Quality of Service (QoS) management. Finally, the third section describes DIA use in multimedia applications and reports on some of the activities in MPEG-21 on extending DIA for the use in the rate distortion modeling for network adaptation. MPEG-21 and rights-information standards conclude the paper.

2 MPEG-21 Standard Organization
The vision of MPEG-21 is to enable transparent and augmented use of multimedia resources across a wide range of network and devices. By creating a full-featured overarching standard this vision can be
achieved and the result is a substantial enhancement in the user’s experience of multimedia. In turn, digital media will grow faster and achieve the dominance that has long been expected [1]. Parts of MPEG-21 standard including titles are presented in Table 1.

Table 1. Parts of MPEG-21 standard

<table>
<thead>
<tr>
<th>PART</th>
<th>TITLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vision, Technologies and Strategy</td>
</tr>
<tr>
<td>2</td>
<td>Digital Item Declaration (DID)</td>
</tr>
<tr>
<td>3</td>
<td>Digital Item Identification (DII)</td>
</tr>
<tr>
<td>4</td>
<td>Intellectual Property Management and Protection Components (IPMP)</td>
</tr>
<tr>
<td>5</td>
<td>Rights Expression Language (REL)</td>
</tr>
<tr>
<td>6</td>
<td>Rights Data Dictionary (RDD)</td>
</tr>
<tr>
<td>7</td>
<td>Digital Item Adaptation (DIA)</td>
</tr>
<tr>
<td>8</td>
<td>MPEG-21 Reference Software</td>
</tr>
<tr>
<td>9</td>
<td>MPEG-21 File Format</td>
</tr>
<tr>
<td>10</td>
<td>Digital Item Processing (DIP)</td>
</tr>
<tr>
<td>11</td>
<td>Persistent Association Technology (PAT)</td>
</tr>
<tr>
<td>12</td>
<td>Test Bed for MPEG-21 Resource Delivery</td>
</tr>
<tr>
<td>13</td>
<td>Empty</td>
</tr>
<tr>
<td>14</td>
<td>Conformance</td>
</tr>
<tr>
<td>15</td>
<td>Event Reporting (ER)</td>
</tr>
<tr>
<td>16</td>
<td>Binary Formant</td>
</tr>
<tr>
<td>17</td>
<td>Fragment Identification of MPEG Resources</td>
</tr>
<tr>
<td>18</td>
<td>Digital Item Streaming (DIS)</td>
</tr>
</tbody>
</table>

For MPEG-21, this is particularly important but is not able to adopt all framework tools. The interrelationships between parts are normatively expressed in many of the parts themselves, particularly those developed later in the standardization. Thus, users who adopt a solution that utilizes a significant subset of parts will generally benefit most from the standard. The complete published standard parts can be found in [2].

3 Digital Item Adaptation

The increase in rich multimedia content is inevitable and has been demonstrated in the last decade. For example, owners of digital cameras already produce tons of images and videos in a relatively short time period. This performs to the diversity of media formats and the variety of Quality of Service (QoS) requirements. In order to cope with this heterogeneity of multimedia content formats, networks and terminals, the development of an interoperable multimedia content adaptation framework has become a key issue. To address this interoperability problem, MPEG has specified a set of normative description tools within the MPEG-21 Digital Item Adaptation (DIA) standard [2, 3] that can be used to support multimedia adaptation. Also, DIA enables the construction of the interoperable adaptation engines that operate independent of the device and coding format. As for device independence, it is guaranteed through a unified description of the environment in which the content is consumed or through which it is accessed.

3.1 Classification of DIA Tools

The DIA tools are divided into categories which are clustered according to their functionalities and use for DIA around the scheme tools and low-level data types. The scheme tools provide uniform root elements for all DIA descriptions as well as some low-level and basic data types, which can be used by several DIA tools independently. The root elements provide a wrapper for complete DIA descriptions, as well as the possibility to generate the so-called DIA description units. A complete DIA description wraps several DIA descriptions (for example, a description of both terminal capabilities and user characteristics). Referencing of external descriptions is also enabled at this level to allow for distributed descriptions. On the other hand, a DIA description unit is used to represent partial information from a complete description within a uniform wrapper element (for example, only the resolution of a terminal). In addition to the root elements, low-level data types are specified to provide commonly used data types such as a set of unsigned integers, single, vector and matrix data types as well as the base, stock function and its possible argument types.

The first category is the Usage Environment Description (UED) tools. It is further subdivided into four major groups: User Characteristics, Terminal Capabilities, Network Characteristics and Natural Environment Characteristics.

User Characteristics define tools related to the user’s content preferences and presentation preferences, as well as accessibility characteristics, mobility characteristics and the destination of a user. The second defines Terminal Capabilities in terms of coding and decoding formats supported by a terminal and associates codec-specific parameters to those formats. Also, device properties such as power and storage characteristics and input-output capabilities (e.g. display capabilities) are defined. The third group of tools comprises Network Characteristics considering the static capabilities of the network (e.g. maximum capacity of a network,
dynamic conditions including the available bandwidth, delay and error performances). The fourth group provides means for describing Natural Environmental characteristics including location and time, and characteristics that pertain to audio-visual aspects.

The Bitstream Syntax Description Link (BSDLink) tool provides a simple referencing mechanism to many of the tools specified within DIA. In particular, reference is made to:

a) tools that are capable of steering the adaptation,
b) the bitstream subject to the resource adaptation,
c) the BSD describing the structure of the bitstream, and
d) several transformations including appropriate parameterization.

Owing to its flexibility and extensibility, this tool enables the design of a rich variety of adaptation architectures and provides the connection between the decision – taking and actual adaptation tools.

The third major category of DIA tools targets the adaptation of media resources in a Digital Item. BSDs enable the adaptation of media resources in an interoperable and format-independent way. Interoperability refers to the possibility of automatically generating descriptions from a bitstream and vice versa. Therefore, Bitstream Syntax Description Language (BSDL), which extends eXtensible Markup Language (XML) Schema Language has been developed. The generic Bitstream Syntax Schema (gBS Schema) enables User’s to contact resource format-independent BSDs, which are also referred to as generic Bitstream Syntax Descriptions (gBSDs).

The fourth category of tools Terminal and Network QoS describes the relationship between constraints, feasible adaptation operations satisfying these constraints and associated qualities. It provides a means to make trade-offs among various adaptation parameters so that an optimal adaptation strategy can be formulated.

The fifth category of tools is referred to as Universal Constraints Description (UCD) tool. It can be used for specifying additional constraints on the provider side as well as on the consumer side of a Digital Item. Constraints can be formulated as limitations or optimization constraints and can be differentiated with regard to individual parts of a resource (for example, a scene of a movie, a group of pictures, or region of interest as a part of an image).

The sixth category of tools, Metadata Adaptability, provides tools for filtering and scaling of XML instances (e.g., MPEG-7 descriptions) with reduced complexity compared to having only the XML instance as well as tools for the integration of several XML instances.

Finally, the seventh category of tools is referred to as Session Mobility, where the configuration-state information that pertains to the consumption of a Digital Item on one device is transferred to the second device. This enables the Digital Item to be consumed on the second device in an adapted way. The concept of DIA tools is summarized in Fig. 1.

**Fig. 1. Categories of DIA tools**

### 3.2 Generic Framework for Adapting the Content

The advantage of scalable multimedia formats is that data are organized in such a way that by retrieving a part of the bitstream, it is possible to render a lower quality version of the content in terms of quality, size, and frame rate. The editing, unlike conventional transcoding, is limited to simple operations such as data truncation. Thus, the content is authored only once, and variety of versions may be obtained on demand from a single bitstream. However, the large variety of competing or complementary coding formats, each featuring its own data structure, hampers of deployment of this strategy since a device providing several content coded in different standards needs as many dedicated software modules as the offered formats to manipulate them.

MPEG-21 DIA defines a generic framework based on XML for adapting the content. This framework is generic in the sense it can be applied to any binary multimedia format and can be used in any device involved in the production, exchange or consumption of the content, including servers, clients or proxies. It should be noted that this tool does not address text-based content such as Hyper...
Text Markup Language (HTML) or Synchronized Multimedia Integration Language (SMIL).

Generally, XML is used for describing the syntax of a bitstream. The resulting document, called Bitstream Syntax Description (BSD) is than transformed (for example, with an Extensible Stylesheet Language Transformation – XSLT style sheet) and used by a generic processor to generate the required adapted bitstream. BSD-based adaptation architecture is depicted in Fig. 5 [3].

![BSD-based adaptation architecture](image)

The first step in thus architecture comprises the BSD generation, which takes the bitstream as input and generates its corresponding BSD. The bitstream and its BSD are subject to actual adaptation. The BSD is transformed according to the usage environment. Subsequently, this transformed BSD is used to generate the adapted bitstream within the bitstream generation process. In practice, the BSD transformation and bitstream generation could be combined for efficiency. This is a simplified view of the architecture, the Bitstream Syntax Schema intervene in the adaptation process.

4 Rate Distortion Modeling for Network Adaptation

The theoretical area that treats data compression from the viewpoint of information theory is called rate distortion theory. In analyzing communication systems that employ data compression, it is becoming commonplace to compare their performances not only with one another but also with absolute bounds provided by rate distortion theory. For example, the basic problem of pattern recognition is related closely to that of data compression. Namely in the pattern recognition problem, several distinct categories of objects are specified a priori. Objects belonging to these categories are than observed in an incomplete, possibly statistical manner, and the problem is to classify them correctly. In particular, rate distortion theory provides knowledge of about how the frequency of faulty categorization will vary with the number and quality of the observation. More importantly, it also gives insight into what set of observation would provide the most information about the objects in question relative to the criterion of proper categorization and, therefore, is of potential value in the design of efficient pattern value in the design of efficient pattern recognition devices. Also, rate distortion theory provides considerable insight into how one can store a large amount of data in a relatively small memory in such a way that it can be accessed as accurately as possible with respect to some specified measure of distortion. Techniques for efficient utilization of memory are applicable to studies in the area of information storage and retrieval.

Part 7 of the MPEG-21 standard defines description tools for multimedia adaptation process based on both the network condition and the available receiver resources. Multimedia compression and streaming have been studied in the rate distortion theory framework that defines tradeoffs between information rate and distortion. Non scalable bitstream switching, adaptive rate scaling, transcoding, scalable coding, distortion-optimized packet scheduling, network adaptive source/channel coding, multiple description coding, etc. have been developed to address real-time adaptation of multimedia content at the server based on the network conditions [4]. In most cases, the network approaches neglect the user experience as well as the capabilities and resource constraints of the receiver (display size, processing power, battery-life, etc.). DIA has defined a set of description tools for adapting multimedia based on the user preferences, terminal capabilities, network performances, and natural environment characteristics [5, 6].

4.1 Previous Works

Research efforts can be classified into three categories:

a) Complexity-scalable multimedia encoding and decoding algorithms,

b) Receiver-driven multimedia streaming based on channel and end-device characteristics,

c) Generic and real complexity matrix and models for multimedia coding and streaming.

A number of authors have considered complexity-scalable coders by focusing on various aspects. Scalable memory complexity reductions compressing I and P frames at the decoder prior to motion-compensated prediction are considered in [7]. An audio decoder with computational scalability has been introduced in [8] by considering
a partial reconstruction of the signal spectrum. Also, decoding of images with scalable complexity has been studied [9]. In the field of adaptive streaming, a receiver driven multicast system that allows receivers to selectively subscribe to multiple channels based on their available bandwidth is proposed in [10]. A generic system for video compression and transmission based on channel and end receiver characteristics in order to accommodate a set of rate, distortion and capability constraints imposed by the receiver is introduced in [11]. A practical implementation of such a system by utilizing a multi-track hinting system, originally introduced in [12] for multimedia adaptation based on complexity was discussed in [13]. A simple rate distortion complexity (R-D-C) adaptation mechanism for wavelet-based video decoding, tacking into account the number of decoded nonzero coefficients used prior to the inverse discrete wavelet transform, was presented in [14], together with the complexity modeling framework.

Rate distortion theory has been extended to complexity-distortion theory and complexity scalability of several simple searching algorithms has been investigated [15]. To enable R-D-C adaptation within the MPEG-21 DIA framework, a practical R-D-C model that relates the various operational R-D points (corresponding to different substreams) to their corresponding decoding complexity is required.

In [16] an abstract receiver referred to as a generic reference machine (GRM) is assumed to be representative of the computation and resource models of the architecture in use. Assuming the GRM as the target receiver, an abstract complexity measure to quantify the decoding complexity of multimedia bitstreams is developed. Given the number of factors that influence the complexity of the receiver, it is impractical to determine at the server side the specific (real) complexity for every possible receiver architecture. Thus, a generic complexity module (GCM) that captures the abstract/GCMs of the employed decoding or streaming algorithm depending on the content characteristics and transmission bit rate is adopted. GCMs are derived by computing the average number of times, the different GRM-operators are executed. A simple GRM supports the set of following operations:

\[ \text{op} = \{ \text{add, multiply, assign} \}. \] (1)

More sophisticated GRMs can be defined to facilitate better mapping of GCMs to architecture-dependent resources (different data and memory types, word lengths, etc.). This involves more complex R-D-C modeling, GCM to real complexity module (RCM) mapping, and bitstream adaptation mechanisms.

4.2 Adaptation Quality of Service Tool in DIA

In DIA, the adaptation QoS tool defines adaptation units (AUs) as a group of video macroblocks, an entire video frame, a certain resolution of frame, a group of pictures (GoP) etc. The GCMs necessary for the decompression and streaming can be transmitted to the receiver at different granularities and for varying size AUs. In general, finer granularity allows better control of the adaptation. This may come at the expense of an increased communication and computational overload. For example, assume that each video GoP is partitioned into \( N \) – independently coded adaptation units. To provide efficient resolution scalability, the maximum size of an AU is usually bounded to be an entire resolution level of a given intra or inter frame in the GoP. Denote the set of AUs that correspond to the decoded resolution and frame rate of a GoP as \( \{ b_1, b_2, \ldots, b_N \} \). Each independent AU \( b_i \), \( 1 \leq i \leq N \), is associated with a set of rate distortion points \( \{ R_i^{(0)}, D_i^{(0)} \} \) with \( j(i) \) indicating the corresponding bitstream adaptation point. An optimization that aims to minimize the overall distortion in the GoP under a rate-constraint \( R_{\text{max}} \) can be stated as

\[ \{ j^{(0)} \}, \lambda^{(0)} \} = \arg \min_{j, \lambda} \left\{ \sum_{i=1}^{N} D_i^{(0)} + \lambda \cdot R_i^{(0)} \right\}, \] (2)

while \( R_{\text{GoP}} \leq R_{\text{max}} \). The Langrangian multiplier \( \lambda \) must be adjusted until the value \( \lambda = \lambda^{*} \) is found where the rate corresponding to the selected points \( j^{*}(i) \) is approximately equal to \( R_{\text{max}} \).

The proposed generic rate distortion complexity model in [16] is able to generate description for DIA. The model addresses network conditions and terminal resources for decoding algorithms. The solution creates a generic model of decoding complexity and adopts it for specific hardware architectures. This technique allows terminals to negotiate with servers and proxies for bitstreams with complexity levels that meet their resource constraints. In that way, this allows delivery to be optimized in an integrated rate distortion complexity setting, thus minimizing distortion under joint rate complexity constraints.

5 Concluding Remarks

MPEG-21 provides a set of standard parts that, in combination with other standards, especially MPEG
While the basic technology Universal Multimedia Access (UMA) is in terms of recent multimedia metadata standards, it is still open whether or not and how these standards will be adopted by industry to create UMA-ready content and applications. However, the end point of multimedia consumption is the user, not the terminal. It is noted that human factors play a significant role in this and are not easily quantified or predicted. There are many factors to be accounted for in an adaptation framework that aims to maximize user experience. User preferences are certainly a dominant factor, but are dynamic in that they are very likely to vary with the content, task and usage environment. Adaptive delivery can be achieved to some expert with tools that are available today. This can be improved with a better understanding of human factors that pertain to the consumption of multimedia.

A rate distortion complexity model can generate metadata necessary for MPEG-21 DIA based on the available receiver resources. The model is generic and is used for other video coders. Also, it could be useful for standardization purposes.

References: