RESEARCH
Open Access

Content on demand video adaptation based on MPEG-21 digital item adaptation

Min Xu[1,4*], Xiangjian He[1], Yu Peng[2], Jesse S Jin[2], Suhuai Luo[2], Liang-Tien Chia[3] and Yusuo Hu[5]

Abstract

One of the major objectives in multimedia research is to provide pervasive access and personalized use of multimedia information. Pervasive access of video data implies the access of cognitive and affective aspects of video content. Personalized use requires the services satisfy individual user's needs on video content. This article attempts to provide a content-on-demand (CoD) video adaptation solution by considering users' preference on cognitive content and affective content for video media in general, sports video and movies in particular. In this article, CoD video adaptation system is developed to support users' decision in selecting their content of interest and adaptively deliver video source by selecting relevant content and dropping frames while considering network conditions. First, video contents are annotated by the description schemes (DSs) provided by MPEG-7 multimedia description schemes (MDSs). Then, to achieve a generic adaptation solution, the adaptation is developed following MPEG-21 Digital Item Adaptation (DIA) framework. We study the MPEG-21 reference software on XML generation and develop our own system for CoD video adaptation in three steps: (1) the content information is parsed from MPEG-7 annotation XML file together with bitstream to generate generic Bitstream Syntax Description (gBSD); (2) Users' preference, network characteristic and adaptation QoS (AQoS) are considered for making adaptation decision; (3) adaptation engine automatically parses adaptation decisions and gBSD to achieve adaptation. Unlike most existing adaptation work, the system adapts the content of interest in the video stream according to users' preference. We implement the above-mentioned MPEG-7 and MPEG-21 standards and provide a generic video adaptation solution. Adaptation based on gBSD avoids complex video computation. Thirty students from various departments were invited to assess the system and their responses have been positive.

1 Introduction

With the explosive growth of video data and the fast development of network technologies, users become accustomed to access video data through network. However, too much video data at the same time might not be suitable for the current situation and the requirement of the users. Users have to spend much time to find video data that they are really interested in. Adapting video content to users' preferences is a key direction for enabling personalized video services. On the other hand, due to the huge size of video data, we're going to need much faster access to the network. However, most of us have limited bandwidth resources. Moreover, users may access and interact with video data on different types of terminals and networks. Personalized video services need to face the

problem of delivering the big size of video data over the network with limited bandwidth in various media environments.

As shown in Figure 1, video adaptation plays an important role between video database and users. It supports exchange, access, and manipulation of multimedia data according to users' preference and network condition. Considering the difference of users' devices, network conditions, and especially their personal preferences of video content, video adaptation systems need personalized access aiming at enhancing the multimedia retrieval process by complementing explicit user preference on video content and user requests with various user's environments including various network conditions and different users' devices.

Video adaptation is a challenging task. Earlier study of encoding reduced video size or provided scalability for video adaptation [1-3]. With the growing number of video formats, attentions were turned towards transcoding video

* Correspondence: min.xu@uts.edu.au
[1]iNEXT, School of Computing and Communications, University of Technology, Sydney, Australia
Full list of author information is available at the end of the article

Figure 1 The role of video adaptation.
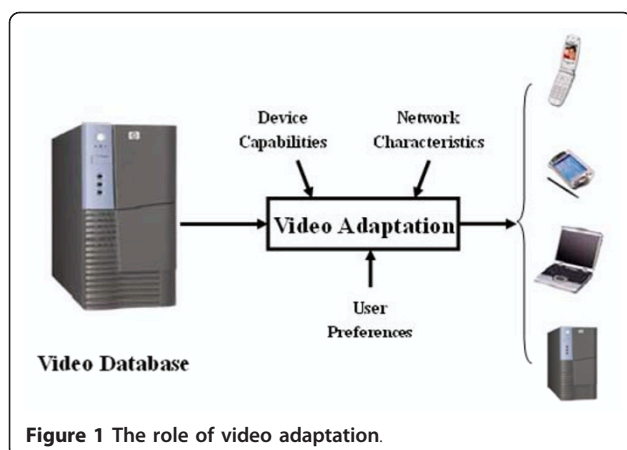
from one format to another. By this way, video became compatible with the new usage environment [4]. Besides encoding and transcoding, another popular adaptation approach is to select, reduce or replace some video elements, such as dropping shots or frames in a video clip [5], dropping pixels and DCT coefficients in an image frame [6] and replacing video sequences with still frames [7]. Although these methods provide feasible ways for video adaptation, there are still some limitations. First, most existing adaptation systems currently focus on achieving a predefined level of visual quality or bitrate without considering users' preference and experience. Second, the current media adaptation solutions tend to be proprietary and hence lack a universal framework. Finally, the methods using transcoding or video elements removal incurs high computational complexity and cost.

In this article, our proposed adaptation system considers video content to adapt videos according to both users' preference and network conditions. The proposed adaptation system improves the existing work by three steps.

First, different from most existing works which focus on achieving a certain defined SNR or bitrate, our adaptation system takes account of user preference and allows users to select video contents of their interests. Sometimes, users may only want to watch video segments of their interests instead of wasting time to browse the whole video. Both affective content and cognitive content are feasible entries for users to access certain video segments. These contents are related to users' understanding and experiencing, which is also a good index to the video. Taking account of users' preferences on video content, the proposed video adaptation allocates more resources to the video parts which attracts users than the unattractive parts.

Second, in order to provide a generic solution to satisfy a wide variety of applications, our system is implemented based on an MPEG-21 digital item adaptation (DIA) framework. Some international standards

such as MPEG-7 and MPEG-21, define the format-independent and environment-independent technologies to support users to exchange, access, consume, trade, and otherwise manipulate digital items in an efficient, transparent and interoperable way [8,9].

Finally, using generic bitstream syntax description (gBSD) which is unaware of bitstream coding format to describe the structure of bitstream provides interoperability in digital item adaptation (DIA). Implementing adaptation based on gBSD instead of the video itself helps to adapt resources quickly with minimal computation cost. It alleviates the computation complexity in transcoding which treats bitstream in a bit-by-bit manner. Furthermore, gBSD can provide structure description at different syntax layer, which enables adaptation at different levels.

The rest of the article is organized as follows. In Section 2, existing video adaptation methods and some related techniques are reviewed. Section 3 briefly introduces our proposed framework and highlights several significant improvements compared with previous study. Sections 4 and 5 introduce the two main components in our framework, i.e., video content analysis and annotation and MPEG-21 digital item adaptation. Section 6 is about the experiments and system evaluation. Finally, discussions and conclusions are in Section 7.

2 Related study
This section reviews video adaptation methods and some related techniques of adaptation.

2.1 Traditional video adaptation
The early study was mostly concerned with network condition for multimedia streaming service. In order to adapt video files for fluctuating network conditions, the network transmission mechanisms [10,11] dynamically adapt video sequence by flexibly dropping portions of elements in a video file, such as enhancement layers, and frames, etc. To make the video scalable for layers or frames dropping, several encoding schemes have been proposed, such as video coding with fine granularity scalability (FGS) [1], multiple description coding (MDC) [2], wavelet-based scalable coding [3], etc. Previous studies focus on how to estimate network quality of services (QoS) and achieve good video quality with limited network resources.

Nowadays, the structure of network is changing from homogeneous to heterogeneous structure. Different network architectures have different capabilities in transmission. Normally, there are two major approaches for dealing with multimedia services via complex heterogeneous network [12]. The first one is adaptive transmission which enforces traditional guaranteed network resource allocation and is tolerant to inevitable fluctuations from various

environments. The problem of maximizing overall quality in adaptive multimedia system has been abstracted to Utility Model [13] to incorporate the dynamics in heterogeneous network environment. Some utility-based adaptation schemes [14-17] have been proposed to optimize the quality of multimedia service with network constraints.

The various capabilities of terminal devices at the end of a network increase the complexity of multimedia services. Some studies [18-20] focus on how to do adaptation concerning limited resource on terminal devices, such as energy, screen size, and presentation capability, etc.

At the same time, the set of emerging rich media formats to be delivered is growing fast. People do not want to bother building specialized adaptation mechanism for every upcoming format. An alternative way to adapt multimedia files between different container formats is transcoding [21,22]. In [19], semantic knowledge about context is used to guide physical adaptation: conversion, scaling and distillation. To handle the bandwidth degradation, some method tries to drop shot or frames in video sequence [5]. Instead of dropping shots completely, some methods retain the keyframes of a shot [7]. Pixels and coefficients are dropped at frame level in [6]. However, the objectives of the adaptation process to save bandwidth utilization cannot satisfy users' requests. Recently, user-specified adaptation has been addressed in literature [23,24]. These study focus on adapting low-level features such as, color depth where users might pay more attention to semantic aspects than low-level features. Semantic video adaptation attracts ever-increasing research efforts [25-27].

2.2 Meta-data-based video adaptation
As the size of the database increases, the traditional data adaptation techniques become insufficient for exploring large amounts of data and finding the desired content. Recently, meta-data based video adaptation attracts more and more research efforts. Before the raw video data can be used to issue queries they must be indexed by content and their indices must be stored as metadata. Metadata is used to facilitate the understanding, characteristics, and management usage of data. The metadata required for effective data management varies with the type of data and context of use.

MPEG-7 defined both syntactic and semantic decompositions to describe syntactic and semantic content in parallel [28,29]. [28] gave an overview of the MPEG-7 description definition language. Besides objectives and specification of MPEG-7 systems, new challenges were discussed in [29], such as the delivery of descriptions either separate or jointly with the audio-visual content, and the like. Recently, media adaptation was achieved by using MPEG-21 digital item adaptation framework [30,31]. In [30], visual content was tailored within the MPEG-21 digital item adaptation (DIA) framework to meet users' visual perception characteristics. In [31], the proposed content on demand method adapted the video with MPEG-21 DIA framework to satisfy users' preference in video content.

2.3 Personalized video adaptation
Personalization is on tailoring a multimedia system to the personal details or characteristics of a user. Due to the difference of users' device, network conditions, and especially their personal preference on video content, adaptation systems need personalized multimedia access aiming at enhancing the multimedia retrieval process by complementing explicit user requests with various user's environments [32-35]. Meanwhile, user preference on video content is vital for achieving personalized video adaptation. [26,31] performed video adaptation by considering video content and user preferences. Most recently, related research focus on adaptation for heterogeneous mobile display devices. In [36], the authors propose a semantic image adaptation scheme to provide mobile users with the most desired image content by integrating the content semantic importance with user preferences under limited mobile display constraints. In [37], the authors attempt to examine several relevant recent developments in ubiquitous media services, especially in the area of content recommendation and user centric content adaptation.

3 A framework of content-on-demand video adaptation with MPEG-21 DIA
Our proposed adaptation system satisfies users' preference on video content which has two primary processes as shown in Figure 2: *content identification and annotation* and *MPEG-21 Digital Item Adaptation*.

First, both cognitive video content and affective video content are identified and later annotated by MPEG-7 structured format. According to users' preference, these video content can be tagged with their own priorities for video adaptation, when necessary.

Second, the content information is parsed from MPEG-7 annotation XML file together with bitstream to generate generic Bitstream Syntax Description (gBSD). When users' request, device capabilities and user preferences is sent to Usage Environment Description (UED), adaptation decision engine determines decision point according to AQoS in order to maximize user satisfaction and adapt it to the constrained environment, such as network condition. The decision point and gBSD will instruct the adaptation operation engine to alter the bitstream and resend to users.

The implementation based on MPEG-21 DIA framework provides a generic adaptation solution. MPEG-21 standard provides some reference softwares to generate and parse XML files describing video source, users'

Figure 2 Adaptation system architecture.
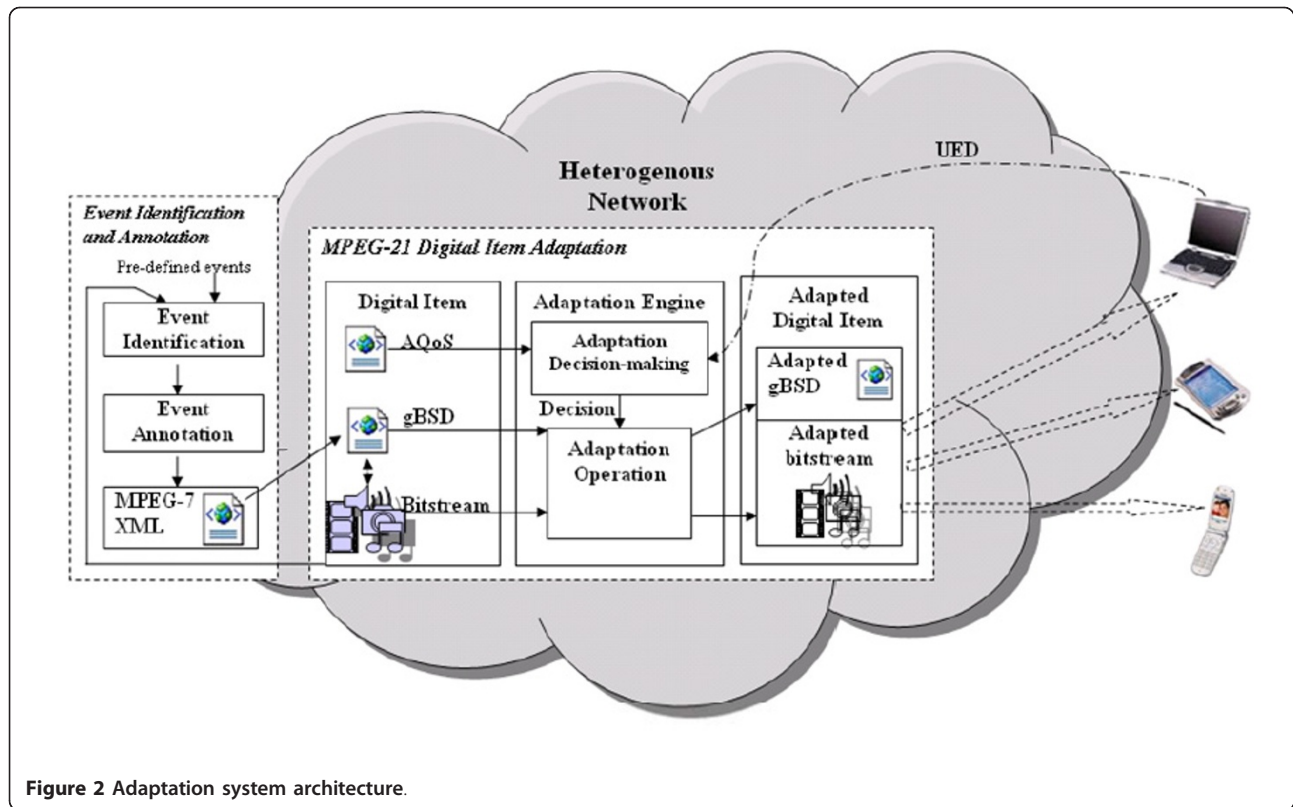
environment, network condition and so on [38]. We use the MPEG-21 reference softwares and develop our system by improving Structured Scalable Meta-formats (SSM) version 2.0 for content agnostic digital item adaptation [39]. The Content on Demand adaptation represents a promising strategy especially for personalized video access applications, where the crucial phase is the video content analysis. In this article, cognitive content in basketball video and affective content in movies are used to demonstrate the performance of the proposed adaptation system. As long as the video content can be annotated, the adaptation scenario is easily extended to other video domains. Compared to [39], the most significant improvement is content on demand adaptation which will be achieved by the following points.

- As a pre-processing step, the content analysis and Annotation module provides an content indexing video to be adapted. The module enables users to directly access preferred content.
- The gBSD structure and descriptions are designed for easy storing and parsing of both bitstream format related information and content related information such as content label and content duration.
- By incorporating users' preference, AQoS is designed to flexibly allocate limited network bandwidth via content.

- According to the changes in average bandwidth, the adaptation decision engine can dynamically make adaptation decision and signal the adaptation operation to adjust the rule of adaptation. Currently, the network bandwidth is estimated by monitoring the transmission time and the file size of past video segment.

4 Video content analysis and annotation
Users' attentions are attracted by the high-level video content such as, what is happening in the video or the video segments which stimulate users' emotional reactions. Capturing and annotating video content can be regarded as a pre-processing step in CoD video adaptation. It provides a feasible way for users to access video content by selecting their interested content.

4.1 Video content analysis
Cognitive content is related to users' understanding of video events which provides a feasible entry for users to access video story. Sometimes, users' preference on video content is not only based on their understanding but also based on their emotional experience beyond video content. Users might prefer "emotional decision" to find their interested video segments because emotional factors directly reflect audiences' attention, evaluation and memory. In this section, content analysis for

both cognitive content and affective content will be briefly presented.

4.1.1 Cognitive content analysis

Cognitive content relies on video domain. Sports videos, possessing an inherent structure due to the special camera technique and the constraints placed on the game through its rules is used to demonstrate CoD adaptation based on cognitive content. Cognitive content analysis is a challenging problem due to the gap between low-level perceptual features and high-level human understanding of videos. We seek some middle-level features, such as specific audio sounds and video scenes. These specific audio sounds have significant clues pointing to interesting events. For example, the sounds of a ball hitting the rim of a basket may be used to confirm the event of a basketball shot being taken. The excited commentator and audience sounds are most likely the consequence of a shot. Additionally, the video scenes provide certain constraints for the event occurrence. By summarizing some heuristic decision rules to combine audio events and video scenes, interesting events are detected. More details can be referred to our previous study [40]. Six basketball events are detected as *Replay, Highlight (goal or shot), Foul, Penalty, Close-up,* and *Normal.*

4.1.2 Affective content analysis

We think of affective contents as those video/audio segments, which may cause audiences' strong reactions or special emotional experiences, such as laughing or fear. Like cognitive content analysis, affective content analysis is also challenging due to the gap between low-level perceptual features and high-level human perception of the media. Emotions are carefully packed and sold with movies. Recently, movie emotion detection occupies a dominant role in multimedia affective computing. In our previous research study [41], a hierarchical affective structure is proposed to analyze affective content hierarchically. First, emotion intensity which describes the degree of agitation is analyzed by using arousal related features. Fuzzy c-mean clustering is used to find three emotion intensity levels based on arousal related features. Sometimes, users might want to watch the content of agitation (high emotion intensity content), where they can not name the detailed emotions. Therefore, emotion intensity levels of *High, Medium,* and *Low* are also analyzed. Second, for each emotion intensity level, detailed emotion types are detected by creating hidden Markov Models with valence related features. Moreover, we propose and experiment feasible solutions other than direct mapping low-level features to high-level affective content. The usage of multiple modalities, such as video, audio, subtitle, etc., are illustrated in [42-44]. Considering users' perception, audio emotional events, dialogue, and emotional words are used as mid-level representation for affective content analysis. In this article, affective content

is categorized as *fear, anger, happiness, sadness,* and *neutral.* The motivation of affective content analysis is to let users directly access their interested video segments.

4.2 MPEG-7-based annotation

Content annotation can be regarded as a pre-processing step in video adaptation, which annotates video segments by using the content analysis results. The video content is annotated by content type and temporal information in order to create tables of content for video documents. It provides a feasible way for users to access video segments by selecting interesting contents. MPEG-7 is a new multimedia standard, designed for describing multimedia content by providing a rich set of standardized descriptors and description schemas. We utilize the description schemes (DSs) of content management and description for temporal information provided by MPEG-7 MDSs to represent the results of content analysis.

- *Temporal Information Description*: The DSs for describing time are based on the ISO 8601 standard, which has also been adopted by the XML schema language. The MediaTime DS describe time information in media streams. Temporal instants and temporal interval is the simplest way to describe MediaTime. A time instant t_1 can be described by a lexical representation using the time point. An interval $[t_1, t_2]$ can be described by its starting point t_1 and a Duration, $t_2 - t_1$.
- *Content Description*: MPEG-7 provides DSs for description of the structure and semantics of Audio-Visual (AV) content. The structural tools describe the structure of the AV content in terms of video segments, frames, still and moving regions, and audio segments. The semantic tools describe the objects, events, and notions from the real world that are captured by the AV content.

In this section, content analysis results in sports video is annotated as an example for cognitive content annotation. A small snippet of content annotation using MPEG-7 XML file is shown in Figure 3. The AudioVisual DS is utilized to describe the temporal decomposition of a video entity. In each TemporalDe-composition DS, some attributes as follows are generated automatically to describe the events.

- MediaTime DS: It specifies the starting point and time intervals of a video segment.
- Event DS: It describes an event, which is a semantic activity that takes place at a particular time or in a particular location.

By using the DSs described above, content analysis results are represented in a standardized and highly

```
  -<Segment xsi:type="AudioVisualSegmentType">
 -<MediaTime>
     <MediaTimePoint>T00:15:46</MediaTimePoint>
     <MediaDuration>T00:01:83</MediaDuration>
   </MediaTime>
 -<Term>
     <termID>Penalty</termID>
   </Term>
     </Segment>
```

Figure 3 An example XML file of MPEG-7 content annotation.

structured format. The MPEG-7 annotation XML files will be parsed to extract content-related information for gBSD generation in the next step.

Similar as cognitive content, affective content are also annotated. Affective content in movies is also annotated by using MPEG-7 DS. The affective content category is used to replace cognitive content name for Term ID in Figure 3.

5 MPEG-21 digital item adaptation (DIA)
Through content on demand video adaptation, a compact representation of the original data can be generated to accommodate view's demands.

The multimedia resource is combined with metadata to describe the network environment, terminal capability

and user characteristic as the fundamental unit of distribution and transaction called the Digital Item (DI). MPEG-21 DIA specifies the syntax and semantics of tools that may be used to assist the adaptation of DI. Different from other adaptation methods, XML file plays an important role in MPEG-21 DIA. In order to provide a generic adaptation for all media types rather than a single format for a specific media type, various network environments, different user characteristics and so on, media data and other information including AQoS, network constraints and users' characteristics are represented by standardized XML files with defined attributes. By parsing these XML files, information which affect adaptation is conveyed between adaptation engine and the media server or media receiver instead of processing the video itself. For XML is a simple explicit language, using XML file ensures the language independence in describing information with different function in adaptation. It also provides freedom in designing adaptation engine to parse and process information from the description in XML files.

The main task of MPEG-21 DIA is actually generating adapted video by selecting video elements in each parcel to meet varying network conditions and maximally satisfy users' preference. Figure 4 shows the MPEG-21 DIA work flow. Two main modules of Adaptation
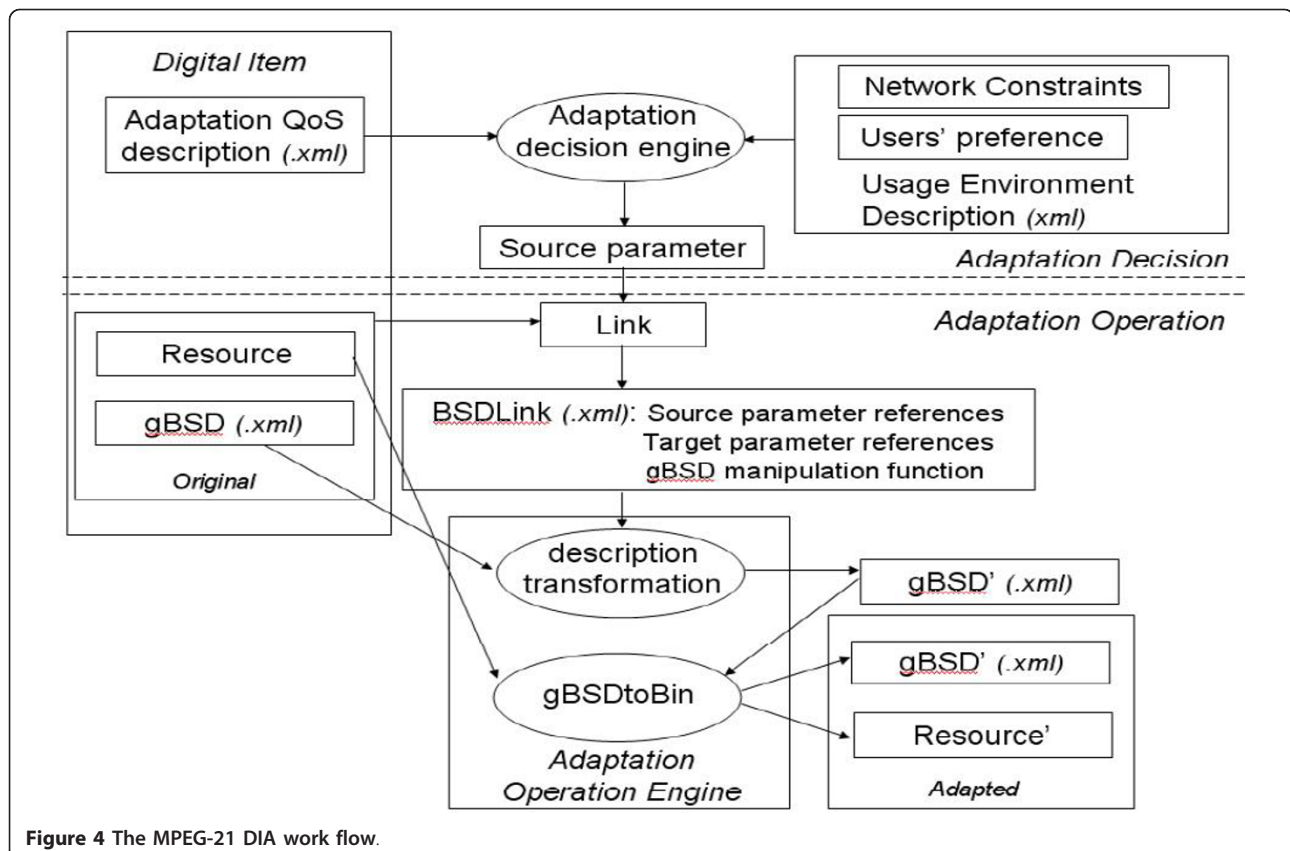
Figure 4 The MPEG-21 DIA work flow.

Decision and Adaptation Operation will be discussed in details as below.

5.1 Adaptation decision

The adaptation decision engine is to make decision of how to adapt each parcel in order to cater for users' preferences and maximize the level of satisfaction under variable network bandwidth. In our system, the input includes Users' preference, Network condition and AQoS. Considering the network constraints and users' preference, adaptation decision engine decides the optimal value of output parameters to instruct frame dropping according to AQoS.

To make an adaptation decision, there are three issues to be considered.

• How to quantify users' preference on video content and provide enough information to help understand preferred content.
• The variety of network conditions is continuous values which needed to be scaled. The network condition exhibits the consecutive changes in values over a period of time, where variables may be predicted from previous value.
• How to define a feasible AQoS of the relationship between possible constrains to allow the selection of optimal adaptation parameters.

In this section, users' preference, network condition, and AQoS will be introduced by considering the above three issues.

5.1.1 User characteristics

Users' characteristics specify general user information, users' preference and usage history. In the proposed system, users are required to input their preferred video content of either affective content or cognitive content, which provides significant cues for later network resources allocation. The user's interface is shown in Figure 5.

Once content of interest are selected through GUI, a usage environment description (UED) XML is generated to store usage related parameters including users preference as shown in Figure 6. Figure 6 shows the user characteristics of accessing interest content for sports
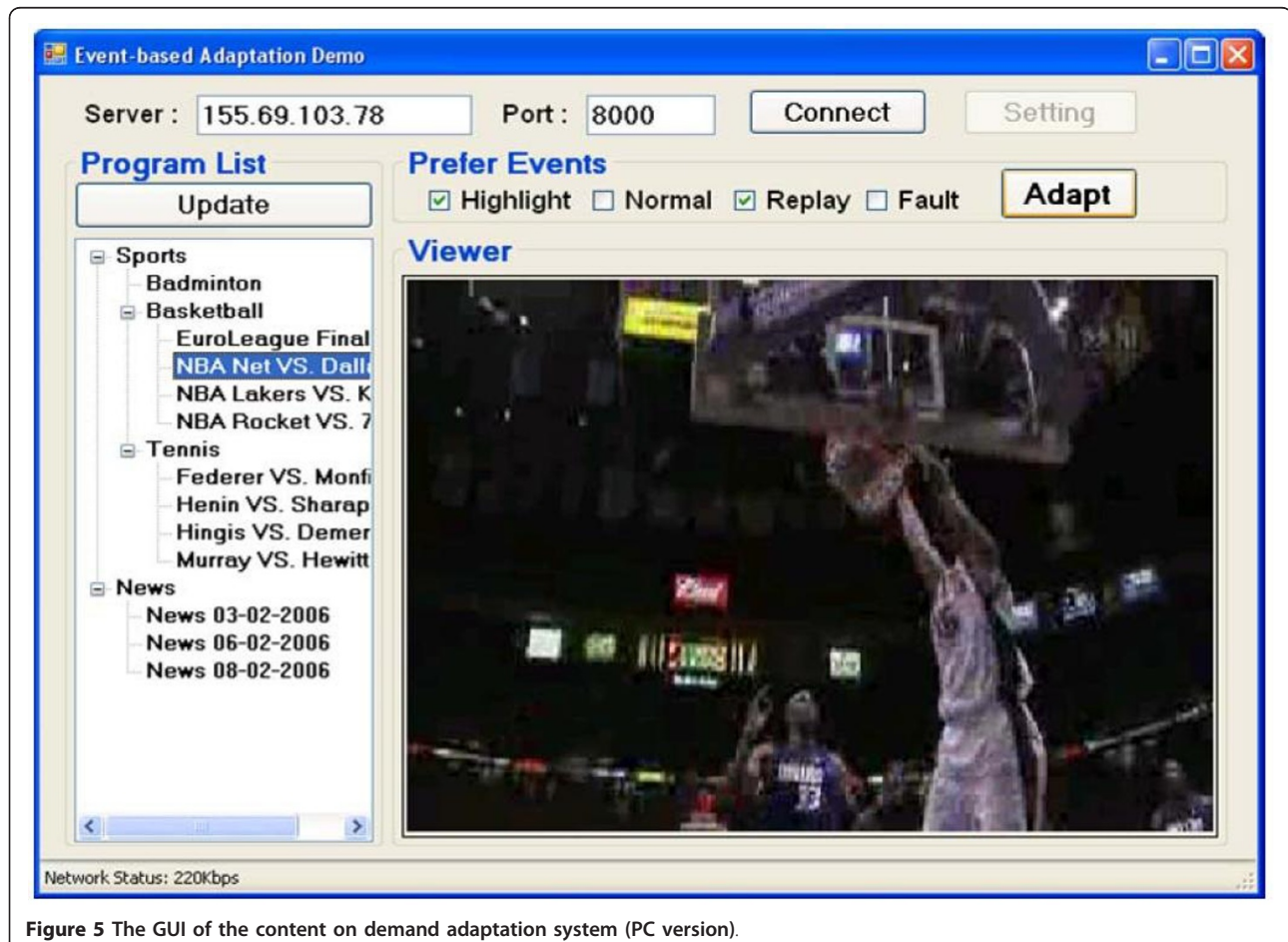

Figure 5 The GUI of the content on demand adaptation system (PC version).

```
<Description xsi:type="UsageEnvironmentType">
  <UsageEnvironment xsi:type="NetworksType">
     <PreferenceEvent highlight="1" normal="0" replay="1" foul="0"
            CloseUp="0" Penalty="1" />
  <Network>
     <NetworkCharacteristic xsi:type="NetworkCapabilityType"
minGuaranteed="100000" maxCapacity="400000" />
     <NetworkCharacteristic xsi:type="NetworkConditionType">
     <AvailableBandwidth average="200000" />
     </NetworkCharacteristic>
  </Network>
 </UsageEnvironment>
</Description>
```

Figure 6 An example of UED on user's characteristics and network condition.

video. Selected contents are marked with "1" while unselected ones are marked with "0".

Besides the user preferred content, how to set priorities on other content is a challenging task. The criteria is maximum information will be provided to the user while minimum resource and time will be occupied. Narrative makes events and contents in a sequence. Cognitive content has coherence of occurrence. Current content rely on the occurrence of previous content or content sequence. From users' point of view, previous content help them to understand the current content. Different from cognitive content, affective content has less coherence and longer duration. Therefore, affective content and cognitive content are considered, respectively.

Quantify users' preference on cognitive content We consider that cognitive contents have some internal causality which can be regarded as context information. For example, *highlight* may occur before *replay*, and *close-up* may appear after *highlight*. From users' point of view, with domain-specific knowledge, the potential context information may help to understand event exactly. For example, we could infer there is a *foul* before *penalty* while the *penalty* after *closeUp* may not be caused due to event *closeUp*. Therefore, the information measure of the previous event provides a feasible way to mark priorities.

To find out the potential causality among various events, we record the times of consecutive events occurrence in the video. Subsequently the information sent from previous event to the current event is measured by calculating the information entropy (IE). To understand current event, previous events are more important than following events. Consequently, to shorten the duration of adapted video, we only consider previous event.

The IE values depend on the video to be adapted. To demonstrate this idea, in the following description, we use a basketball video that lasts 1 hour and includes 38

close-up, 45 *normal*, 40 *highlight*, 6 *relay*, 5 *penalty*, and 5 *foul*. Table 1 shows the frequencies of consecutive events occurrence.

The event sequence can be represented by discrete random sequence e_t, where $t = 1,2,...,n$. The possible value of each e_t is from event set E which consists of six kinds of events $\{E_1, E_2, E_3, E_4, E_5, E_6\}$ which are $\{close\text{-}up, normal, highlight, replay, penalty, and foul\}$. The probability $P(F_i) = n_{E_i}/n$, where n_{E_i} is the number of E_i take place in e_t. The given constraint is

$$\sum_{i=1}^{6} P(E_i) = 1. \tag{1}$$

From Table 1, the probability $P(E_i/E_j)$ could be calculated, giving the constraint

$$\sum_{i=1}^{6} P(E_i/E_j) = 1. \tag{2}$$

where i and j are used to indicate different event. The information sent from e_j to e_i can be calculated by

$$I_{e_{j \to i}} = \log_2 \left(\frac{p(e_i/e_j)}{P(e_i)} \right). \tag{3}$$

Once users select their interested event, the priority of previous event will be set according to these information measurement values shown in Table 2.

Figure 7 shows examples of event priorities setting. To help understanding, the corresponding values in Figure 7 are highlighted in Table 2. Looking at the second row in Figure 7, when *highlight* is selected, the previous event (*closeUp*) is labeled as the value of information measurement sent from *closeUp* to *Highlight*, i.e., -3.474 according to Table 2. Sometimes, if an event gets more than one priorities from different contexts, the higher value is chosen as its priority label. A peculiar example is shown in the third row of Figure 7. The selected events are *replay* and *penalty*. As the event before

Table 1 The frequencies of consecutive events occurrences

Number of Occurrence		Current event					
		CloseUp	Normal	Highlight	Replay	Penalty	Foul
	CloseUp		0	2	27	3	3	3
	Normal		30	0	10	2	1	1
Next	Highlight		1	38	0	0	1	0
Event	Replay		4	1	1	0	0	1
	Penalty		3	1	0	1	0	0
	Foul		0	3	2	0	0	0

Table 2 The measurement of information sent from previous event to users' preferred event

	Number of occurrence	Current event					
	CloseUp	Normal	Highlight	Replay	Penalty	Foul	
	CloseUp	N.A.	-2.635	1.306	0.874	1.136	1.136
	Normal	1.284	N.A.	-0.373	0.040	-0.697	-0.697
Next event	Highlight	-3.474	1.552	N.A.	N.A.	-0.527	N.A.
	Replay	1.288	-0.966	-0.783	N.A.	N.A.	2.218
	Penalty	1.134	-0.711	N.A.	2.224	N.A.	N.A.
	Foul	N.A.	0.896	0.474	N.A.	N.A.	N.A.

penalty, *replay* get the value of 2.224 (from Table 2). However, as selected event, the priority of *replay* can be "selected" which is higher than the priority labeled with 2.224. Therefore, the *Replay* is labeled as "selected" in stead of 2.224.

Quantify users' preference on affective content Users can access affective content by inputting their preferred affective intensity level (high, medium, and low) or detailed affective types. Cognitive intensity is a subjective and consecutive concept. Normally, users prefer the affective content with either high intensity or low intensity, since video highlights are most likely conjoin with the content with high affective intensity and low affective intensity. The criteria of priority setting is to not only provide users as much as the content with their preferred affective intensity but also make users understand the whole video story. Fuzzy clustering was used for affective intensity detection [45]. It is based on minimization of the objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2, 1 \le m < \infty \qquad (4)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by

Original sequence
    …CloseUp → Highlight → Normal → Highlight → Foul → Replay → Penalty…
Highlight selected
    …CloseUp(-3.474)→ Highlight(selected)→ Normal(1.552)→ Highlight (selected)…
Replay & Penalty selected
    …Foul(2.218) → Replay(selected or 2.224) → Penalty(selected)…

Figure 7 Examples of events' priorities setting.

$$u_{ij} = \cfrac{1}{\sum_{k=1}^{C} \left(\cfrac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}}, \qquad (5)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad (6)$$

This iteration will stop when $\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}(k) \right| \right\} < \varepsilon$, where ϵ is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

After the processing of fuzzy clustering, we get center matrix, fuzzy partition matrix and the values of the objective function during iterations. Fuzzy partition matrix is actually membership function matrix, which indicate the degree to which the data points belong to the different clusters. Once user select the content with certain affective intensity level, the corresponding cluster is actually selected. The degree of each data point belong to the corresponding data cluster can be calculated from membership function matrix. The degree shows the distance from each point to the center of the cluster, which can be used for setting the priorities of allocating limited network resources. The closer data point should get more resources because it is more likely to have the same affective level as user's preference. On the other hand, Affective types have no much correspondence among each other. If users select their preferred affective type, "selected" and "unselected" will be directly set to the content sequence.

5.1.2 Network condition
Network is the medium for multimedia transmission. The quality of network condition has major effect on multimedia service quality. Heterogeneous network structure requires the transmission of multimedia files adapted to fluctuating network condition in order to achieve good multimedia service quality, while saving network resources and preventing network congestion. In our study, the network condition is measured by bandwidth. Figure 6 records the initial network condition which is described by network minimum and maximum capability and the average available bandwidth. Since the device may from various network, such as LAN, MANET, Internet, wireless network, cellular network, etc., the initial values are set by the profile of current access network. whereafter, these values are updated by constant measurement of the network condition. A monitor can be set at the server side to survey the past network condition. The monitor detects the transmission time of previous fixed-size segment of adapted media file to compute the bandwidth available in the network.

Since network variety is continuous, the attributes of current network capability during the negotiation period can be estimated by past network condition, which is supported by MPEG-21 standard for describing both static and time-varying network conditions.

5.1.3 Adaptation QoS and decision making

The AQoS specifies the relationship between constraints, feasible adaptation operations satisfying these constraints and possibly associated utilities or qualities. In our case, there are two constraints (content priority and network condition) and one feasible operation (drop segment or frame) designed as follows:

- Besides the highest priority of "selected", the value of priorities for cognitive content are from -5 to 5. The events with priorities below 0 are regarded as no contribution to understanding the following event. Therefore, all priorities value less than 0, are scaled as the lowest one. For priority values, the priorities are mapped to five scales: 5 (selected); 4 (above 2); 3 (1 ~ 2); 2 (0 ~ 1); 1 (below 0). For affective content, users can access their interested content by selecting either affective intensity level or affective types. If users select a certain intensity level, each video content has a value of the membership function to the selected intensity level. The value of membership function are from 0 to 1, which are averagely scaled into 5. 5(0.8 1); 4(0.6 0.8); 3(0.4 0.6); 2(0.2 0.4); 1(0 0.2). On the other hand, if users select a specific emotion type, the content with selected type is set to 5 while the rest are set to 1.
- According to the changes in average bandwidth, the adaptation decision engine dynamically changes adaptation decision and signals the adaptation operation to change the rule of adaptation. We divide the network bandwidth into five scales: 1 (below 50 kbps); 2 (50 kbps ~ 100 kbps); 3 (100 kbps ~ 200 kbps); 4 (200 kbps ~ 300 kbps); 5 (above 300 kbps). The adaptation decision is made by considering the value of the previous network scales.
- In our case, adaptation operation is to drop different portion of videos. Three scales of operation are: 0 (drop the whole video segment with I, P, B frames); 1 (drop P, B frames); 2 (drop B frames); 3 (remain as the original video segment with no frame dropped).

Table 3 shows an example of feasible AQoS. The AQoS is made by considering three rules:

- Content selection is based on user preference.
    - For cognitive content, according to users' preferred events and the information measurement for the previous events, the parcels containing highly preferred events or high information entropy are most likely to remain after adaptation.

Table 3 An example of feasible AQoS

Frame dropping Scales		Network condition scales				
		1	2	3	4	5
Content priority scales	1	0	0	0	1	1
	2	0	0	1	1	2
	3	0	1	1	2	2
	4	1	1	2	2	3
	5	2	2	2	3	3

- For affective content, video segments belonging to users' preferred affective type remain as much as possible after adaptation. Video content which is more likely to be users' preferred affective intensity level is allocated more resources after adaptation.
- According to the current bandwidth, adaptation keeps as many frames as possible to convey the original story.
- With the bandwidth changing, the user's preferred segments (including selected content and related content which help to understand selected events) have higher priorities of retaining all types of frames.

According to the current content priority and the current network condition, decision engine find the optimal adaptation scheme from AQoS to not only satisfy all constraints but also maximize or minimize optimization value. In order to best utilize the network resources, the AQoS is dynamically updated as the multimedia service time pass by. The parameters in Table 3 is dynamically changing based on the events distribution in the rest of media file.

5.1.4 The output source parameters

XML files which contain adaptation parameters are the output of adaptation decision as shown in Figure 8. The adaptation unit in decision file makes rule for adaptation operation of corresponding content. Here, cognitive content are used to demonstrate the output source parameters. There is no much difference for affective content. Two snippets of the output decision files for the same adaptation unit under different network condition (220 k versus 80 k) are shown in Figure 8. Users' selected content (event) is *Highlight*. CONTENT ID from 1 to 6 orderly correspond to content (events) *Highlight, Normal, Replay, Foul, CloseUp, Penalty* (Similar as shown in Figure 9). CONTENT_PRIORITIES is dedicated selected event and the priority for other events according to correlation of selected event (Table 2). Later, these priorities and network bandwidth is mapped to CONTENT_PRIORITIES_SCALE and BANDWIDTH_SCALE (see, Section 5.1.3). Considering BANDWIDTH_SCALE and EVENT_PRIORITIES_S-CALE, the decision engine output adaptation decision of how to drop frames by TEMPORAL_LAYERS. As

Figure 8 Examples of adaptation decisions.

described in Section 5.1.3, higher layer number indicates more frame remaining (i.e., less frame dropping). The adaptation unit with TEMPORAL_LAYER equals to 0 will be totally deleted in adapted video (see the last illustrated AdaptationUnit.) The decision file is changing dynamically according to network condition. The BANDWIDTH indicates the current bandwidth which is listed and used for decision making in the next AdaptationUnit. The two examples show the difference in TEMPORAL_LAYERS when network bandwidth is degraded from around 220 kbps to 80 kbps. The change indicates fewer frames are pertained in the adaptation because of network gradation.

5.2 Adaptation operation

Adaptation operation is conducted based on adaptation decision. To achieve a format independent adaptation, the adaptation operation is performed on video description. Later on, the adapted video description guides adapted video production. In this section, we will focus on video description generation and later adaptation operation.

5.2.1 Generic bitstream syntax description (gBSD)

To generate an appropriate video description which contains the information of video format and interesting events is an important and necessary step for further adaptation. The gBSD is an important element of DI, which allows the adaptation of multimedia resources by a single, media resource-agnostic processor. An XML description of the media resource's bitstream syntax can

be transformed to reflect the desired adaptation and then be used to generate an adapted version of the bitstream. In our system, BSDL and gBS Schema [46] are used for parsing a bitstream to generate its gBSD description.

The bitstream is described based on parcels. In our case, each parcel corresponds to a video segment with certain content label. The content and duration related information is extracted from the MPEG-7 XML annotation file which has been introduced in Section 4.2. Considering events have ranks according to various users' preference, we introduce so-call Content-Level to mark different events for users to access their events of interest. Figure 9. shows how the interaction between users defined events and metadata can be inserted in gBSD. Furthermore, frame dropping is a feasible way to adapt to the variation of network situation. We introduce Temporal-Level 0, 1, 2 to mark I-frame, P-frame and B-frame in gBSD. An example of gBSD is shown in Figure 10.

5.2.2 Operation

Transformation instruction directs the operation how to use the parameters in decision file to do adaptation on gBSD and resource file.

• *Description transformation*: The original gBSD file is used to describe the structure of the media file. Transformation instruction sets a mechanism on how to use adaptation decision parameters to alter the structure of media file. It initiates the engine to retain, delete or update gBSD units based on decision. Comparing temporal level of every frame with
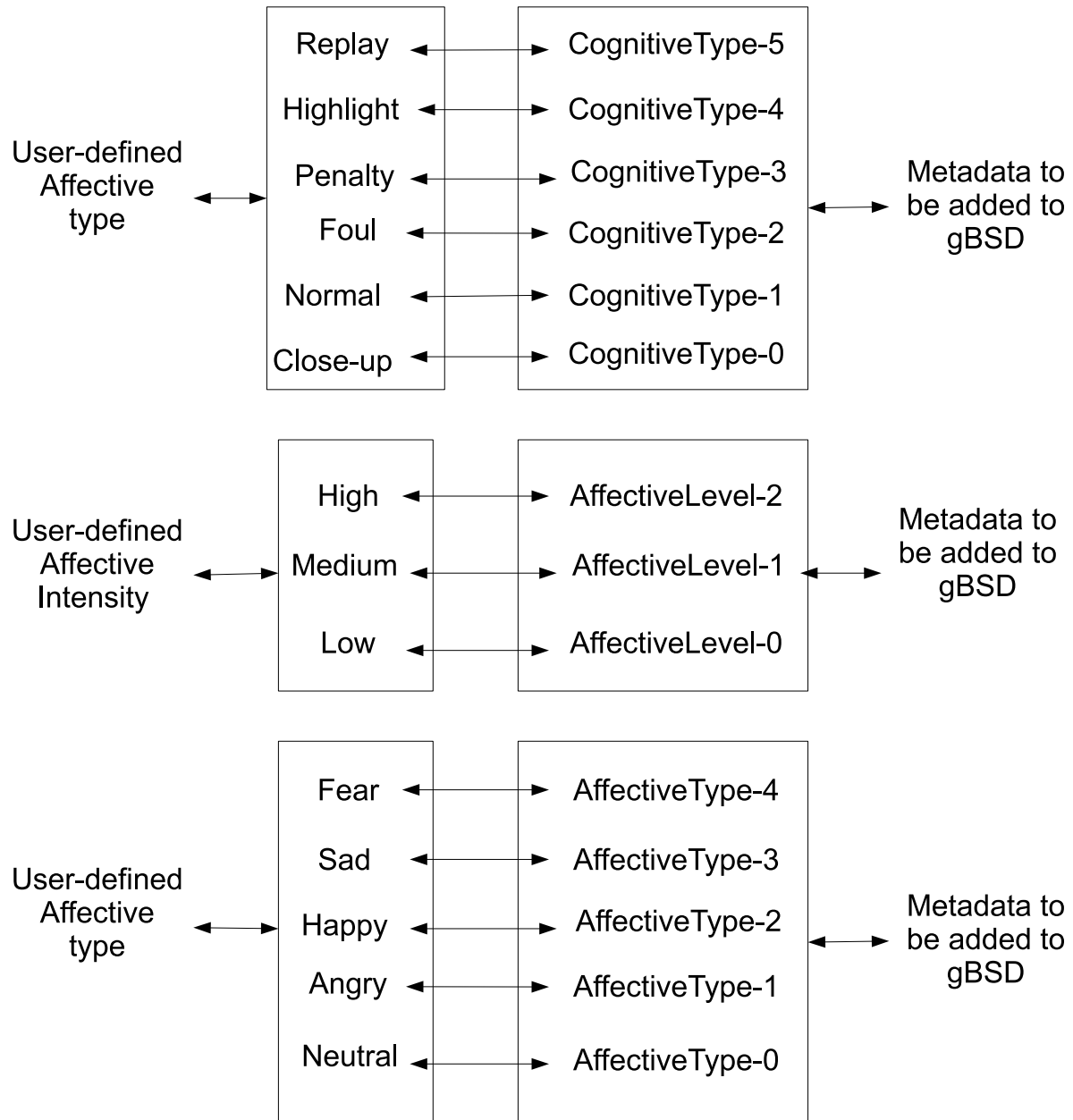
Figure 9 The interaction between users and metadata.

corresponding decision parameter, the adaptation operation engine decides whether to drop or retain certain units in gBSD. The adapted structure can be used in proxies to do multi-step adaptation. It can also be used to do further operations based on analyzed content.

• *gBSDtoBin*: This part generates the final adapted media file based on the adapted structure from Description transformation. It parses the description of target adapted file to understand the structure of it. Based on the altered structure of gBSD specified in original media file context, the gBSDtoBin can select, drop or change certain frames with the help of the indication in altered structure. The adapted media file comprises the selected snippets from original resource.

```
 –<dia:Description xsi:type="gBSDType" id="basketball_gBSD"
bs1:bitstreamURI="basketball.mp4">
    <gBSDUnit syntacticalLabel=":M4V:VOL" start="0" length="19" />
 –<gBSDUnit start="19" length="324083" marker="Content-3">
        <gBSDUnit syntacticalLabel=":M4V:I_VOP" start="19" length="6158"
marker="Temporal-0" />
        <gBSDUnit syntacticalLabel=":M4V:P_VOP" start="6177" length="1301"
marker="Temporal-1" />
        <gBSDUnit syntacticalLabel=":M4V:B_VOP" start="7478" length="328"
marker="Temporal-2" />
        ......  .....  ......
```

Figure 10 An example of gBSD.

In our streaming adaptation scheme, we need to operate Description Transformation and gBSDtoBin at the same time in order to provide real-time adaptation. The network monitor will detect network condition and adjust the network attribute in UED file. If the change in network leads to a change in decision file, operation in gBSD and resource file will be suspended. Adaptation operation engine reparse the adaptation rules from decision file. After being invoked from pending, the following adaptation operation in gBSD and resource file will be based on the new rules.

5.2.3 Adapted video
Since the adaptation is based on video description, video in various formats can be adapted. In the original video sequence, every gBSD unit (video segment) is associated with a content related label to indicate the associated content (event, affective content type or affective intensity level). Once the user selects his preferred content, the priorities of frame retaining are assigned to every gBSD unit according to content' priority level and current bandwidth estimation. An example of adapted video

is shown in Figure 11. Here, cognitive content is used as an example. The video sequence is adapted based on the content' priorities as shown in Figure 7. AQoS associates retaining priority in the order of event sequence is: 1, 5, 3, 5, N.A., N.A., N.A.. A lower priority means it is likely to be partially or totally dropped when network condition degrades. Based on network bandwidth constraint, adaptation decision engine eliminates unqualified decisions that cause excessive network bandwidth. The final decision is the one utilizing most of the reserved bandwidth while preserving most of interesting events.

6 Experiments and system evaluation
An content-driven adaptation system is implemented for the MPEG-21 digital item adaptation framework. This section is focused on the experiments for system evaluation.

The experiment was conducted on a testbed which is composed of our streaming adaptation server connected to a laptop. The server is a PC with single CPU Intel Pentium 4 2.8 GHz, 1.5 GB RAM, running Microsoft Windows XP Professional operating system. The laptop is with Intel Pentium-M 1.86 GHz CPU, 1 GB RAM, running Windows XP Professional, which is connected to the server with IEEE 802.11b wireless card. For the source video, we loop an MPEG-4 video stream where I frames appear 2.5 times per second, P frames 12.5 times per second, and B frames 15 times per second with average bit rate 300 kbps.

Our system is tested under three different network conditions with bandwidth at 80, 150, and 220 kbps, respectively.

Figure 11 Structure of original and adapted video sequence.

Figure 12 shows the adaptation result when user selecting "Highlight" and when bandwidth are 250 kbps or 150 kbps. Figure 12 illustrates the adapted quality at different events periods in terms of bit rate. According to the Table 2, "Highlight" has the priority of five. "Normal" event is taken as event related to "Highlight" and set its Event Priority to three. It is obviously that our adaptation system provides user's preferred event at higher quality than the bandwidth resource could afford. This is done by reserving bandwidth and saving streaming time from unimportant event. After the unimportant events are streamed in a shorter time slot than the time they play-back, the saved time slot could be utilized to stream important events at higher quality.

The traditional adaptation methods which ignore users' preference on video content are evaluated by PSNR. Unfortunately, PSNR may not be able to achieve a reasonable evaluation for CoD adaptation. The users' preference is an important role on the proposed CoD adaptation system, which is a subject concept depends on individual's understanding and perception. In this case, a user study is carried our for 30 students who are selected from both engineering and non-engineering departments. Through the user study we evaluate users' satisfaction with the following three cases.

- Without considering uses' preferred content, the video is adapted to satisfy the variation in network conditions.
- Assuming static network condition, the adaptation only takes account of content on demand (CoD).
- CoD video adaptation on time-varying network conditions.

We adopted the double stimulus impairment scale (DSIS) [7] method with some modifications to evaluate our adaptation system. The 30 students are requested to come to our lab and attend the user evaluation

individually. Each of them is asked to watch the adapted video first. And then, the original video is shown to them. The reason of asking them to watch the adapted videos before the original videos is to avoid semantic impression from original video affecting their evaluations. After watching the original videos, the students have full understandings on the video stories. They are able to evaluate their understanding based on watching the adapted videos only. Three groups of experiments are conducted to evaluate the system performance.

6.1 Frame dropping only based on network conditions

The purpose for this group of experiments is to investigate users' comments on our scheme of frame dropping. All the content priorities are set to scale 5. With the variation of network bandwidth, the adaptation engine drops frames. The students are asked to compare the adapted video to the original video and vote on video quality and satisfaction of semantic conveying. Table 4 shows the voting result. From Table 4, compared to video quality, users are not satisfied with the adapted video on whether the adapted video can properly convey semantic meanings. A possible reason may be that when network condition degrades badly, the adaptation only can retain I frames. Video composed of only I frames lose large amount of information so that there is hardly any semantics left.

6.2 Adaptation considering CoD only

This group of experiments evaluates whether users are satisfied with the adapted result by content dropping. Assuming of static network conditions, we asked students to input their content preference. The preferred content can be more than one. Based on their inputs, adaptation is performed according to AQoS in Table 3 when setting constant scale four to network condition. Five scales are provided for their voting on satisfaction of semantic conveying. The voting result (Table 5) shows
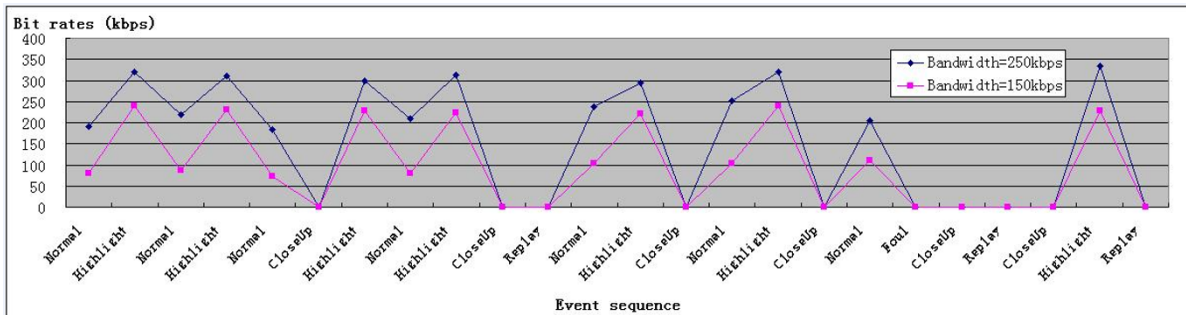

Figure 12 Curves of adapted video quality.

Table 4 Case 1: user voting on degradable network condition

	Bad (%)	Poor (%)	Fair (%)	Good (%)	Excellent (%)
Video quality	3.3	6.7	30.3	53.3	3.3
Semantic conveying	6.7	16.7	40	30.3	3.3

that most of the students are satisfied with the content-driven adaptation as it provides them their preferred content in the limited bandwidth.

6.3 CoD video adaptation on time-varying network condition

In this section, experiments are designed to test user's acceptance of the adapted video stream by frame dropping. Each user compares adapted video with original video and gives a evaluation on semantic understanding for the adapted video clip based on the five scales from "Bad" to "Excellent", corresponding to semantic quality from "ambiguous" to "complete understanding" respectively. Frames are partially dropped based on defined priority and available network bandwidth. We introduce three adapted versions with 220 kbps, 150 kbps and 80 kbps for transmitting. Table 6 shows the voting result of the three adapted video streams. Obviously, network degradation affects the user's understanding of the video. However, the high priority assigned to retaining user's preferred content has resulted in an adapted video that is still able to retain and convey the preferred information. For small drop in bandwidth, there is only a marginal effect on user's perception (i.e., semantic quality) of the adapted video. Comparing Case 3 (Table 6) to Case 1 (Table 4), we find that CoD adaptation greatly improves users' satisfaction on semantic quality even if the bandwidth declines to 80 kbps.

7 Discussions and conclusions

In this article, a robust content on demand adaptation system is achieved to support users preference when accessing video under time-variety network condition. The proposed content on demand video adaptation system follows MPEG-21 digital item adaptation framework, which provides a generic platform for the interaction between users and multimedia database. Content selection and frame dropping are effective and efficient ways to meet users' preference and adapt to the variation of network

Table 5 Case 2: user voting on content-on-demand adaptation

	Bad (%)	Poor (%)	Fair (%)	Good (%)	Excellent (%)
Voted result	0.0	10.0	23.3	36.7	30.0

Table 6 Case 3: user voting on content-on-demand adaptation over degradable network

	Bad (%)	Poor (%)	Fair (%)	Good (%)	Excellent (%)
BW = 220 kbps	0.0	6.7	16.7	46.6	30.0
BW = 150 kbps	3.3	13.3	30.0	36.7	16.7
BW = 80 kbps	6.7	13.3	33.3	33.3	13.3

condition. MPEG-21 digital item adaptation helps to reduce computational complexity through XML manipulations. In order to provide a generic adaptation for all media types rather than a single format for a specific media type, standardized XML files with defined attributes are used to represent various network environments, different user characteristics, media data and other information including AQoS, network constraints and user characteristics. Moreover, adaptation based on XML files instead of directly on the video itself helps to minimize the computation cost. It alleviates the computation complexity in video transcoding, which treats bitstream in a bit-by-bit manner.

Author details
[1]iNEXT, School of Computing and Communications, University of Technology, Sydney, Australia [2]School of Design, Communication and IT, University of Newcastle, Newcastle, Australia [3]School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore [4]National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China [5]Google Inc., Mountain view, CA, USA

Competing interests
The authors declare that they have no competing interests.

References
1. W Li, Overview of fine granularity scalability in mpeg-4 video standard. IEEE Trans. Circ. Syst. Video Technol. **11**(3), 301–317 (2001). doi:10.1109/76.911157
2. VK Goyal, Multiple description coding: Compression meets the network. IEEE Signal Process. Mag. **18**(5), 74–93 (2001). doi:10.1109/79.952806
3. K Shen, E Delp, Wavelet based rate scalable video compression. IEEE Trans. Circ. Syst. Video Technol. **9**, 109–122 (1999). doi:10.1109/76.744279
4. J Xin, CW Lin, MT Sun, Digital video transcoding. IEEE Proc. **93**, 84–97 (2005)
5. KT Fung, YL Chan, Siu WC, New architecture for dynamic frame-skipping transcoder. IEEE Trans. Image Process. **11**(8), 886–900 (2002). doi:10.1109/TIP.2002.800890
6. S Benyaminovich, O Hadar, E Kaminsky, Optimal transrating via dct coefficients modification and dropping, in *Proceedings of the 3rd Conference on Information Technology: Research and Education*, vol. 3. (Hsinchu, Taiwan, 2005), pp. 100–104
7. SF Chang, D Zhong, R Kumar, Real-time content-based adaptive streaming of sports video, in *IEEE CVPR Conference on IEEE Workshop Content-Based Access to Video/Image Library*, (Kauai, HI, USA, 2001), pp. 139–146
8. BS Manjunath, P Salembier, T Sikora, Introduction to mpeg-7: Multimedia content description interface, (Wiley, 2002) ISBN: 978-0-471-48678-7

9. Mpeg-21 digital item adaptation, ISO/IEC Final Standard Draft ISO/IEC 21000-7:2004(E), ISO/IEC JTC 1/SC 29/WG 11/N5895. (2004)
10. M Naghshineh, M Willebeek-LeMair, End-to-end qos provisioning in multimedia wireless/mobile networks using an adaptive framework. IEEE Commun. Mag. **35**(11), 72–81 (1997). doi:10.1109/35.634764
11. Z Wang, C Jan, Quality of service routing for supporting multimedia applications. IEEE J. Sel. Area Commun. **14**(7), 1228–1234 (1996). doi:10.1109/49.536364
12. J Gecsei, Adaptation in distributed multimedia systems. IEEE Multimedia **4**(2), 58–66 (1997). doi:10.1109/93.591164
13. S Khan, KF Li, EG Manning, The utility model for adaptive multimedia systems, in *International Conference on Multimedia Modeling*, (Singapore, 1997), pp. 111–126
14. JG Kim, Y Wang, SF Chang, Content-adaptive utility-based video adaptation, in *International Conference on Multimedia and Expo*, vol. 3. (Baltimore, Maryland, 2003), pp. 281–284
15. X Wang, H Schulzrinne, An integrated resource negotiation, pricing, and qos adaptation framework for multimedia applications. IEEE J Sel. Areas Commun. **18**(12), 2514–2529 (2000). doi:10.1109/49.898734
16. JT Park, JW Baek, JWK Hong, Management of service level agreements for multimedia internet service using a utility model. IEEE Commun. Mag. **39**(5), 100–106 (2001). doi:10.1109/35.920863
17. Y Wang, JG Kin, SF Chang, Utility-based video adaptation for universal multimedia access (UMA) and content-based utility function prediction for real-time video transcoding. IEEE Trans. Multimedia. **9**(2), 213–220 (2007)
18. W Yuan, K Nahrstedta, SV Advea, DL Jonesb, RH Kravetsa, Design and evaluation of a cross-layer adaptation framework for mobile multimedia systems, in *Proceeding of SPIE-the International Society for Optical Engineering*, vol. 5019. (Santa Clara, California, USA, 2003), pp. 1–13
19. M Metso, A Koivisto, J Sauvola, A content model for the mobile adaptation of multimedia information. J. VLSI Signal Process. **29**, 115–128 (2001). doi:10.1023/A:1011131816588
20. WH Cheng, CW Wang, JL Wu, Video adaptation for small display based on content recomposition. IEEE Trans. Circ. Syst. Video Technol. **17**, 43–58 (2007)
21. R Han, P Bhagwat, R Lamaire, T Mummert, V Perret, J Rubas, Dynamic adaptation in an image transcoding proxy for mobile web browsing. IEEE Personal Commun. **5**(6), 9–17 (1998)
22. A Garcia, H Kalva, B Furht, A study of transcoding on cloud environments for video content delivery, in *Proceedings of ACM multimedia workshop on Mobile cloud media computing*, (Firenze, Italy, 2010), pp. 13–18
23. D Cotroneo, G Paolillo, C Pirro, S Russo, A user-driven adaptation strategy for mobile video streaming applications, in *Proceedings of 25th IEEE International Conference on Distributed Computing Systems Workshops*, (Columbus, Ohio, USA, 2005), pp. 338–344
24. M Hicks, A Nagarajan, R van Renesse, User-speciified adaptive scheduling in a streaming media network, in *Proceedings of International Conference on Multimedia and Expo*, vol. 3. (Baltimore, Maryland, 2003), pp. 281–284
25. F Pereira, PV Beek, AC Kot, J Ostermann, Special issue on analysis and understanding for video adaptation. IEEE Trans. Circ. Syst. Video Technol. **15**(10), 1197–1199 (2005)
26. M Bertini, R Cucchiara, A Bimbo, A Prati, Semantic adaptation of sport videos with user-centred performance analysis. IEEE Trans. Multimedia. **8**(3), 433–443 (2006)
27. H Huang, X Zhang, Z Xu, Semantic Video Adaptation using a Preprocessing Method for Mobile Environment, in *proceedings of International Conference on Computer and Information Technology (CIT)*, (Bradford, UK, 2010), pp. 2806–2810
28. J Hunter, An overview of the MPEG-7 description definition language (DDL). IEEE Trans. Circ. Syst. Video Technol. **11**(6), 765–772 (2001). doi:10.1109/76.927438
29. O Avaro, P Salembier, MPEG-7 Systems: overview. IEEE Trans. Circ. Syst. Video Technol. **11**(6), 760–764 (2001). doi:10.1109/76.927437
30. J Nam, YM Ro, Y Huh, M Kim, Visual content adaptation according to user perception characteristics. IEEE Trans. Multimedia. **7**(3), 435–445 (2005)
31. M Xu, J Li, LT Chia, JS Jin, Y Hu, BS Lee, D Rajanet, Event on demand with MPEG-21 video adaptation system, in *Proceedings of the ACM Multimedia Conference*, (Santa Barbara, CA, USA, 2006), pp. 921–930
32. S Acharya, B Smith, P Parnes, Characterizing user access to videos on the world wide web, in *ACM/SPIE Multimedia Computing and Networking*, vol. 3969. (San Jose, CA, USA, 2000), pp. 130–141
33. S Ramanathan, P Rangan, Architectures for personalized multimedia. IEEE Multimedia. **1**, 37–46 (1994)
34. W Lee, J Wang, A user-centered remote control system for personalized multimedia channel selection. IEEE Trans. Consumer Electron. **50**(4), 1009–1015 (2004). doi:10.1109/TCE.2004.1362492
35. H Harroud, M Ahmed, A Karmouch, Policy-driven personalized multimedia services for mobile users. IEEE Trans. Mobile Comput. **2**, 16–24 (2003). doi:10.1109/TMC.2003.1195148
36. W Yin, J Luo, CW Chen, Event-based semantic image adaptation for user-centric mobile display devices. ACM Trans. Multimedia Comput. Commun. Appl. **13**(3), 432–442 (2011)
37. W Yin, X Zhu, CW Chen, Contemporary ubiquitous media services: Content recommendation and adaptation, in *Proceedings of PerCom Workshops*, (Seattle, USA, 2011), pp. 129–134
38. ISO/IEC Final Standard Draft ISO/IEC 21000-8:2005(E), ISO/IEC JTC 1/SC 29/WG 11/N5895: Information technology - multimedia framework (mpeg-21)-part 8: Reference software (2005)
39. D Mukherjee, G Kuo, A Said, Structured scalable meta-formats (ssm) version 2.0 for content agnostic digital item adaptation - principles and complete syntax, (Imaging Systems Laboratory, HP Laboratories Palo Alto, HPL-2003-71, 2003)
40. M Xu, CS Xu, LY Duan, JS Jin, S Luo, Audio keywords generation for sports video analysis. ACM Trans. Multimedia Comput. Commun. Appl. **4**(2), 11 (2008)
41. M Xu, JS Jin, S Luo, LY Duan, Hierarchical movie affective content analysis based on arousal and valence features, in *Proceedings of the ACM Multimedia Conference*, (Vancouver, British Columbia, Canada, 2008), pp. 677–680
42. M Xu, LT Chia, JS Jin, Affective content analysis in comedy and horror videos by audio emotional event detection, in *Proceedings of IEEE International Conference on Multimedia & Expo*, vol. 61. (Amsterdam, The Netherlands, 2005), pp. 2–5
43. M Xu, S Luo, JS Jin, T Liu, Using dialogue to detect emotion segments in movies, in *Proceedings of Asia-Pacific Workshop On Visual Information Processing (VIP)*, (Tainan, Taiwan, 2007), pp. 54–58
44. M Xu, LT Chia, H Yi, D Rajan, Affective content detection in sitcom using subtitle and audio, in *Proceedings of the 12th International MultiMedia Modelling Conference (MMM)*, (Beijing, China, 2006), pp. C1–C6
45. M Xu, S Luo, JS Jin, Affective content detection by using timing features and fuzzy clustering, in *Proceedings of IEEE Pacific Rim Conference on Multimedia*, vol. 5353/2008. (Tainan, Taiwan, 2008), pp. 1685–692
46. G Panis, A Hutter, J Heuer, H Hellwagner, H Kosch, C Timmerer, S Devillers, M Amielh, Bitstream Syntax Description: A Tool for Multimedia Resource Adaptation Within MPEG-21. Signal Process: Image Commun. **18**(8), 721–747 (2003). doi:10.1016/S0923-5965(03)00061-4