Event on Demand with MPEG-21 Video Adaptation System

Min Xu[1,2] Jiaming Li[1] Liang-Tien Chia[1] Jesse S. Jin[2]
Yiqun Hu[1] Bu-Sung Lee[1] Deepu Rajan[1]
[1]School of Computer Engineering Nanyang Technological University,
Singapore, 639798 Email: {mxu, liji0006, asltchia, y030070, ebslee, asdrajan}@ntu.edu.sg
[2]School of Design, Communication and IT, University of Newcastle
Callaghan 2308, Australia Email: {M.Xu}@studentmail.newcastle.edu.au {Jesse.Jin}@newcastle.edu.au

ABSTRACT

In this paper, we present an event-on-demand (EoD) video adaptation system. The proposed system supports users in deciding their events of interest and considers network conditions to adapt video source by event selection and frame dropping.

Firstly, events are detected by audio/video analysis and annotated by the description schemes (DSs) provided by MPEG-7 Multimedia Description Schemes (MDSs). And then, to achieve a generic adaptation solution, the adaptation is developed following MPEG-21 Digital Item Adaptation (DIA) framework. We look at early release of the MPEG-21 Reference Software on XML generation and develop our own system for EoD video adaptation in three steps: 1) the event 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 video of events with interest according to users' preference. Implementation following MPEG-7 and MPEG-21 standards provides a generic video adaptation solution. gBSD based adaptation avoids complex video computation. 30 students from various departments were invited to test the system and their responses has been positive.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*Distributed applications*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Algorithms, Design, Experimentation

Keywords

Event on Demand, Adaptation, Event detection, Annotation, MPEG-7, MPEG-21

1. INTRODUCTION

With the increasing amount of multimedia data and the development of multimedia communication techniques, there is an increasing need to develop effective and efficient video adaptation systems. In various media environments, users may access and interact with multimedia content on different types of terminals and networks. As shown in Fig. 1., video adaptation play an important role between video database and users. It supports exchange, access, and manipulation of multimedia data according to users' preference and network condition.

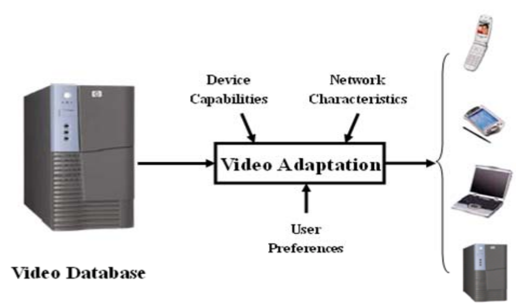

Figure 1: The role of video adaptation

Video adaptation is still a challenging field. Earlier work developed encoding schemes to reduce video size or provide the scalability for video adaptation [14, 9, 22]. With the increasing amount of video formats, attention turned towards transcoding video from one format to another in order to make the video compatible with the new usage environment [25]. Besides encoding and transcoding, a popular adaptation approach is to select, reduce or replace some video elements, such as dropping shots and frames in a video clip [7], dropping pixels and DCT coefficients in an image frame[4], replacing video sequences with still frames [5] etc. Although these methods provide feasible ways for video adaptation, there are still some limitations as follows.

1. Most existing adaptation systems currently focus on achieving a certain defined SNR or bitrate without considering users' preference and experience.

2. The current media adaptation solutions tend to be proprietary and therefore lack a universal framework.

3. The transcoding and video elements removal will incur high computational complexity and cost.

In this paper, our proposed system alleviates the above limitations by three steps mentioned below.

Firstly, our adaptation system takes account of users' preference by asking users to select events in which they are interested. Sometimes, users may only want to watch those interesting video segments instead of wasting time browsing the whole video. Event is a feasible entry for users to access certain video segments because events are related to user's understanding, which is also a good semantic index to the video. Taking account of user's preference on events, Video adaptation allocates more resources to the video parts which attracts users than the unattractive parts.

Secondly, 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 technologies to support users to exchange, access, consume, trade and otherwise manipulate digital items in an efficient, transparent and interoperable way [16, 1].

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 this paper is organized as follows. Adaptation system architecture will be briefly introduced in Section 2. Section 3 will review our previous work on sports event detection. Detected interested events are annotated by MPEG-7 descriptors which are presented in Section 4. Section 5 presents the details of Digital Item Adaptation (DIA), which is the core of our proposed adaptation system. Experiments and results are presented in Section 6. Some related conclusions are in Section 7.

2. RELATED WORK

This section presents traditional adaptation methods and some related techniques of adaptation.

The early work was mostly concerned with network condition for multimedia streaming service. In order to adapt video files for fluctuating networks, the network transmission mechanisms [19, 24] dynamically adapt video sequence by flexibly dropping portions of elements in video file, such as enhancement layers, frames and so on. To make the video scalable for layers or frames dropping, several encoding schemes have been proposed, such as Fine Granularity Scalability (FGS) video coding [14], Multiple Description Coding (MDC) [9], wavelet-based scalable coding [22], etc. Previous works focus on how to estimate network 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 approaches when dealing with problems in multimedia services via complex heterogeneous network [8]: adaptive transmission which enforces traditional guaranteed network resource allocation, and adaptive applications which are more tolerant to inevitable fluctuations in the supporting environment's performance. The problem of maximizing overall quality in adaptive multimedia system has been abstracted to Utility Model [12] to incorporate the dynamics in heterogeneous network environment. Some utility-based adaptation schemes [13, 23, 21] have been proposed to optimize the quality of multimedia service under network constraints.

The various capabilities of terminal devices at the end of a network increase the complexity of multimedia services. Some works [26, 17] focus on how to do adaptation concerning limited resource on terminal devices, such as energy, screen size, 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 [3, 10]. In [17], semantic knowledge about context is used to guide physical adaptation: conversion, scaling and distillation. To handle the bandwidth degradation, One method tries to drop shot or frames in video sequence [7]. Instead of dropping shots completely, some methods retain the keyframes of the shot [5]. Pixels and coefficients are dropped at frame level [4]. However, objectives of the adaptation process to reduce the bandwidth utilization cannot satisfy users' requests. User-specified adaptation has been addressed in literature [6] [11]recently. These work focus on adapting low-level features such as, color depth whereas users might pay more attention to semantic aspects than low-level features.

Our proposed adaptation system bridges the gap between users' preference on semantic and video content. 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' environment, network condition and so on [2]. We use the MPEG-21 reference softwares and develop our system by improving Structured Scalable Metaformats (SSM) version 2.0 for content agnostic digital item adaptation [18].

3. THE ADAPTATION SYSTEM

Before we describe our system step by step, an overview of system architecture will be explained. Our proposed adaptation system has two primary processes (Fig. 2): *event identification and annotation* and *MPEG-21 Digital Item Adaptation*. The MPEG-21 DIA can be employed on server and any intermediate proxies between server and client, on the basis of the user-provided preferences. In the experiments, the adaptation process is performed on the server side and will be easily extended to adaptation on proxies.

Firstly, some pre-defined interesting events are identified by our event identification and annotation module and stored in MPEG-7 structured format. Detecting semantic highlights or events in video has attracted much interests. Most of the previous methods rely on a single feature (audio, video or transcripts) and each feature provides some hints on interesting video events or video highlights. Recently, our

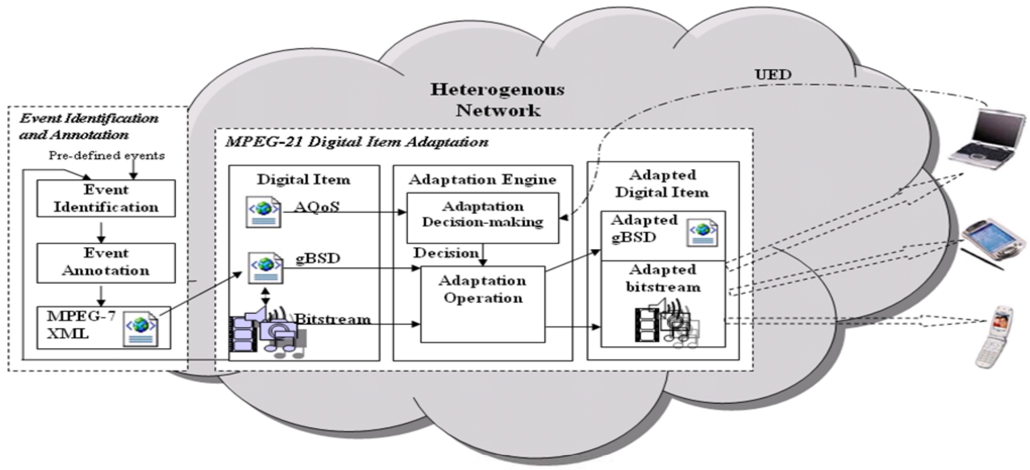

Figure 2: Adaptation system architecture

event detection work based on audio sounds identification and video scene detection has shown promising results [15]. According to users' preference, these detected events can be tagged with their own priorities for video adaptation, when necessary.

Secondly, the event 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 Event on Demand adaptation represents a promising strategy especially for streaming applications, where the crucial phase is the video content analysis. The event identification and annotation engine is more suited for video in which the events are easily detected such as sports videos. However the adaptation scenario is easily extended to other video domains. for example, the movie can be adapted by user selected scenes, and news video can be adapted by news content, etc. In this paper, video scenarios from basketball games which are presented in previous publication [15] are used to demonstrate how well the proposed adaptation system performs. Compared to [18], the most significant improvement is EoD adaptation which will be achieved by the following points.

- As a pre-processing step, The Event Identification and Annotation module provides an event indexing video to be adapted. The module enables users to directly access preferred events.
- The gBSD structure and descriptions are designed for easy storing and parsing of both bitstream format related information and event related information such as event label and event duration.
- By incorporating users' preference, AQoS is designed to flexibly allocate limited network bandwidth via events.
- 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. EVENT DETECTION AND MPEG-7 BASED ANNOTATION

Capturing interesting events can be regarded as a pre-processing step in video adaptation. It provides a feasible way for users to access video content by selecting interesting events.

4.1 Event Identification

Event identification 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 hints pointing to interesting events. For example, the sound of a ball hitting the rim of the 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 work [15]. Six basketball events are detected as: *Replay*, *Highlight* (goal or shot), *Foul*, *Penalty*, *Close-up* and *Normal*.

4.2 Event Annotation

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 provided by MPEG-7 MDSs to represent the results of event identification. A small snippet of event annotation using MPEG-7 XML file is listed in Fig. 3. The AudioVisual DS is utilized to describe the temporal decomposition of a video entity. In each TemporalDecomposition

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.

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

Figure 3: An example XML file of MPEG-7 event annotation

By using the DSs described above, event detection results are represented in a standardized and highly structured format. The MPEG-7 annotation XML files will be parsed to extract event-related information for gBSD generation in the next step.

5. MPEG-21 DIGITAL ITEM ADAPTATION (DIA)

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 files 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. Fig. 4. shows the MPEG-21 DIA work flow. Two main modules of Adaptation Decision and Adaptation Operation will be discussed in detail below.

5.1 Adaptation Decision

The adaptation decision engine is to make decision of how to adapt each parcel in order to cater for user preferences and maximize the level of satisfaction under variable network bandwidth. In our system, the input includes Users'

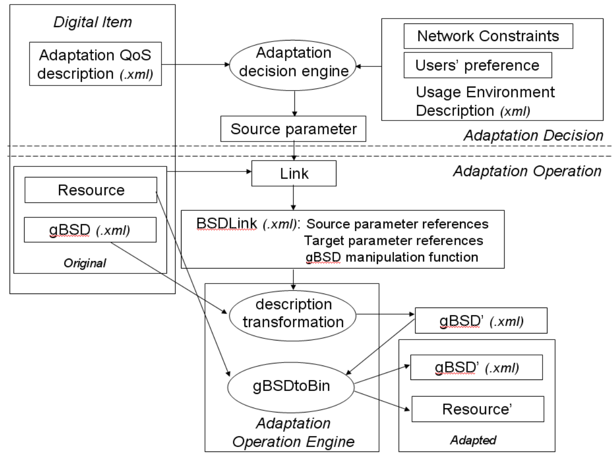

Figure 4: The MPEG-21 DIA work flow

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.

1. How to quantify users' preference on event and provide enough information to help understand preferred events.
2. 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.
3. How to define a feasible AQoS of the relationship between possible constrains to allow the selection of optimal adaptation parameters.

5.1.1 User Characteristics

Users' characteristics specify general user information, users' preference and usage history etc. In the proposed system, users are required to input their preferred events which provides significant cues for later event priority setting. The user's interface is shown in Fig. 6 and Fig. 7.

Once events 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 Fig. 5. Selected events are marked "1" while unselected ones are marked "0".

Besides the user preferred event, how to set priorities on other events is a challenging task. The criteria is maximum information will be provided to the user while minimum resource and time will be occupied. We consider that events 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

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

Figure 5: An example of UED on User's Characteristics and Network Condition

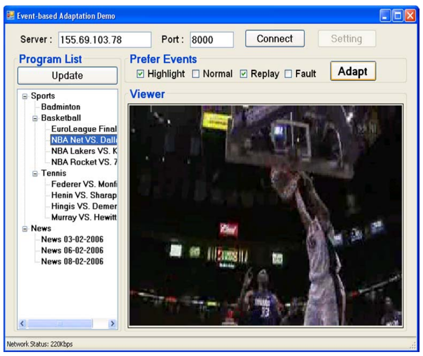

Figure 6: The GUI of the event on demand adaptation system (PC Version)

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 to be adapted by parsing MPEG-7 annotation file. Subsequently the information sent from previous event to the current event is measured by calculating the information entropy (IE). We note that in sport knowledge domain, the previous event provides more cues than the following events. Consequently, to shorten the duration of adapted video, we only consider the 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 6 kinds of events $\{E_1, E_2, E_3, E_4, E_5, E_6\}$ which are {*Close-up, Normal, Highlight, Replay, Penalty and Foul*}. The probability $P(E_i) = n_{E_i}/n$, where n_{E_i} is the number of $E_i$take place in e_t. The given constraint is

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

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

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

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

$$I_{e_{j->i}} = log_2(\frac{p(e_i/e_j)}{P(e_i)}). \quad (3)$$

Once users select their interested event, the priority of previous event will be set according to these information measurement values shown in Table 2. Fig. 8 shows examples of

Figure 7: The GUI of the event on demand adaptation system (PDA Version)

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 8: Examples of events' priorities setting

event priorities setting. To help understanding, the corresponding values in Fig. 8 are highlighted in Table 2. Looking at the second row in Fig. 8, when *Highlight* is selected, the previous event (*CloseUp*) is labelled as the value of the measurement of information 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 Fig. 8. The selected events are *Replay* and *Penalty*. as the event before *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 labelled with 2.224. Therefore, the *Replay* is labelled as "selected" in stead of 2.224.

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 work, the network condition is measured by bandwidth. Fig. 5 records the initial

Table 1: The frequencies of consecutive events occurrences

Number of Occurence		Current Event					
		CloseUp	Normal	Hightlight	Replay	Penalty	Foul
Next Event	CloseUp	0	2	27	3	3	3
	Normal	30	0	10	2	1	1
	Highlight	1	38	0	0	1	0
	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 Occurence		Current Event					
		CloseUp	Normal	Hightlight	Replay	Penalty	Foul
Next Event	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
	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.

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. We set a monitor at server side to survey 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 (event priority and network condition) and one feasible operation (drop segment or frame) designed as follows:

1. Besides the highest priority of "selected", the value of priorities 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. Therefore, the priorities are mapped to 5 scales: 5 (selected); 4 (above 2); 3 ($1 \sim 2$); 2 ($0 \sim 1$); 1 (below 0).

2. 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 5 scales: 1 (below 50Kbps); 2 ($50Kbps \sim 100Kbps$); 3 ($100Kbps \sim 200Kbps$); 4 ($200Kbps \sim 300Kbps$); 5 (above 300Kbps). The adaptation decision is made by considering the value of the previous network scales.

Table 3: An example of feasible AQoS

Frame Dropping Scales		Network Condition Scales				
		1	2	3	4	5
Event 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

3. 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 four rules:

1. Event selection is based on user preference. 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.

2. According to the current bandwidth, adaptation keeps as many frames as possible to convey the original story.

3. With the bandwidth changing, the user's preferred segments (including selected events and previous events which help to understand selected events) have higher priorities of retaining all types of frames.

According to the current event 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 Fig. 9. The adaptation unit in decision file makes rule for adaptation operation of corresponding event. Two snippets of the output decision files for the same adaptation unit under different network condition (220K v.s. 80k) are shown in Fig. 9. Users' selected event is *Highlight*. EVENT_ID from 1 to 6 orderly correspond to events *Highlight*, *Normal*, *Replay*, *Foul*, *CloseUp*, *Penalty* (Similar as shown in Fig. 10). EVENT_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 EVENT_PRIORITIES_SCALE and BANDWIDTH_SCALE (See 5.1.3). Considering BANDWIDTH_SCALE and EVENT_PRIORITIES_SCALE, the decision engine output adaptation decision of how to drop frames by TEMPORAL_LAYERS. As described in 5.1.3, higher layer number indicates more frame remaining. 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 220Kbps to 80Kbps. 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 generic bitstream syntax description (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 [20] 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 event. The event and duration related information is extracted from the MPEG-7 XML annotation file which has been introduced in 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. Fig. 10. 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 as Fig. 11.

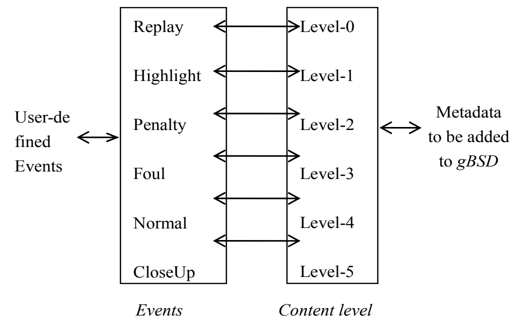

Figure 10: The interaction between users and metadata

```
-<dia:Description xsi:type="gBSDType" id="basketball_gBSD"
bs1:bitstreamURI="basketball.mpg4">
   <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 11: An example of gBSD

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.

1. *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 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 analysis content.

2. *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.

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

```
<Decision>
  <AdaptationUnit>
    <IOPin Id="EVENT_ID">2</IOPin>
    <IOPin Id="EVENT_PRIORITIES">1.552</IOPin>
    <IOPin Id="EVENT_PRIORITIES_SCALE">3</IOPin>
    <IOPin Id="BANDWIDTH">226017</IOPin>
    <IOPin Id="BANDWIDTH_SCALE">4</IOPin>
    <IOPin Id="TEMPORAL_LAYERS">2</IOPin>
  </AdaptationUnit>
  <AdaptationUnit>
    <IOPin Id="EVENT_ID">1</IOPin>
    <IOPin Id="EVENT_PRIORITIES">Selected</IOPin>
    <IOPin Id="EVENT_PRIORITIES_SCALE">5</IOPin>
    <IOPin Id="BANDWIDTH">226017</IOPin>
    <IOPin Id="BANDWIDTH_SCALE">4</IOPin>
    <IOPin Id="TEMPORAL_LAYERS">3</IOPin>
  </AdaptationUnit>
  <AdaptationUnit>
    <IOPin Id="EVENT_ID">5</IOPin>
    <IOPin Id="EVENT_PRIORITIES">-3.474</IOPin>
    <IOPin Id="EVENT_PRIORITIES_SCALE">0</IOPin>
    <IOPin Id="BANDWIDTH">226017</IOPin>
    <IOPin Id="BANDWIDTH_SCALE">4</IOPin>
    <IOPin Id="TEMPORAL_LAYERS">0</IOPin>
  </AdaptationUnit>

  ...
</Decision>
```

```
<Decision>
  <AdaptationUnit>
    <IOPin Id="EVENT_ID">2</IOPin>
    <IOPin Id="EVENT_OCCURRENCE">1.552</IOPin>
    <IOPin Id="EVENT_PRIORITIES">3</IOPin>
    <IOPin Id="BANDWIDTH">82516</IOPin>
    <IOPin Id="BANDWIDTH_SCALE">2</IOPin>
    <IOPin Id="TEMPORAL_LAYERS">1</IOPin>
  </AdaptationUnit>
  <AdaptationUnit>
    <IOPin Id="EVENT_ID">1</IOPin>
    <IOPin Id="EVENT_PRIORITIES">Selected</IOPin>
    <IOPin Id="EVENT_PRIORITIES_SCALE">5</IOPin>
    <IOPin Id="BANDWIDTH">82516</IOPin>
    <IOPin Id="BANDWIDTH_SCALE">2</IOPin>
    <IOPin Id="TEMPORAL_LAYERS">2</IOPin>
  </AdaptationUnit>
  <AdaptationUnit>
    <IOPin Id="EVENT_ID">5</IOPin>
    <IOPin Id="EVENT_PRIORITIES">-3.474</IOPin>
    <IOPin Id="EVENT_PRIORITIES_SCALE">0</IOPin>
    <IOPin Id="BANDWIDTH">82516</IOPin>
    <IOPin Id="BANDWIDTH_SCALE">2</IOPin>
    <IOPin Id="TEMPORAL_LAYERS">0</IOPin>
  </AdaptationUnit>

  ...
</Decision>
```

Figure 9: Examples of adaptation decisions

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.

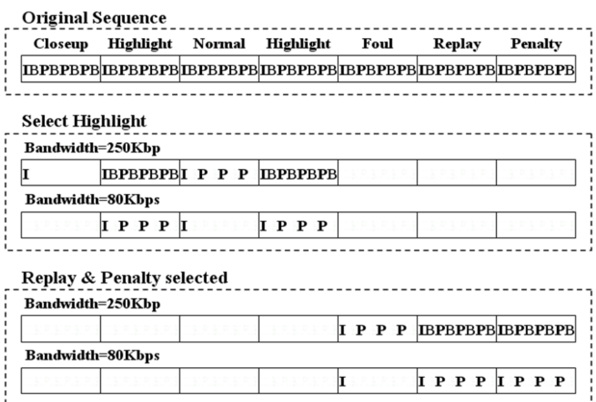

Figure 12: Structure of original and adapted video sequence

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 *Content-Level* to indicate the associated event. Once the user selects his preferred events, the priorities of frame retaining are assigned to every gBSD unit according to events' priority level and current bandwidth estimation. An example of adapted video is shown in Fig. 12. The video sequence is adapted based on the events' priorities as shown in Fig. 8. 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 event-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 and a PDA. The server is a PC with single CPU Intel Pentium 4 2.8GHz, 1.5GB of RAM, running Microsoft Windows XP Professional operating system. The laptop is with Intel Pentium-M 1.86GHz CPU, 1GB RAM, running Windows XP Professional, which is connected to the server with IEEE 802.11b wireless card.The PDA we use is O2 Xda Atom PDA, with Windows Mobile 5.0 platform, 128M memory, Intel 416MHz CPU, connecting to network through Wi-Fi. 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 300kpbs. The 15 students use PDA and others use the laptop. Our system is tested under three different network conditions with bandwidth at 80, 150 and 220 Kbps, respectively.

Fig. 13 shows the adaptation result when user selecting "Highlight" and when bandwidth are 250kpbs or 150kpbs. Fig. 13 illustrates the adapted quality at different events periods in terms of bit rate. According to the Table 2, "Highlight" has the priority of 5. "Normal" event is taken as event related to "Highlight" and set its Event Priority to 3. 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

Table 4: Case 1: User voting on degradable network condition

	Bad	Poor	Fair	Good	Exellent
Video quality	3.3%	6.7%	30.3%	53.3%	3.3%
Semantic conveying	6.7%	16.7%	40%	30.3%	3.3%

and saving streaming time from unimportant event. After the unimportant events are streamed in a shorter time slot than the time they playback, 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 EoD adaptation. The users' preference is an important role on the proposed EoD 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.

1. Without considering uses' preferred event, the video is adapted to satisfy the variation in network conditions.

2. Assuming static network condition, the adaptation only takes account of event on demand (EoD).

3. EOD video adaptation on time-varying network conditions.

We adopted the double stimulus impairment scale (DSIS) [5] method with some modifications to evaluate our adaptation system. Users view the adapted video first, followed by the original video in order to avoid semantic impression from original video affecting evaluation of adapted video understanding. 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 event 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 EoD only

This group of experiments evaluates whether users are satisfied with the adapted result by event dropping. Assuming of static network conditions, we asked students to input their event preference. The preferred event can be more than one. Based on their inputs, adaptation is performed according to AQoS in Table 3 when setting constant

Table 5: Case 2: User voting on event-on-demand adaptation

	Bad	Poor	Fair	Good	Exellent
Voted result	0.0%	10.0%	23.3%	36.7%	30.0%

Table 6: Case 3: User voting on event-on-demand adaptation over degradable network condition

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

scale 4 to network condition. 5 scales are provided for their voting on satisfaction of semantic conveying. The voting result (Table 5) shows that most of the students are satisfied with the event-driven adaptation as it provides them their preferred events in the limited bandwidth.

6.3 EoD 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 5 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 3 adapted versions with 220Kbps, 150Kbps and 80Kbps for transmitting. Table 6 shows the voting result of the 3 adapted video streams. Obviously, network degradation affects the user's understanding of the video. However, the high priority assigned to retaining user's preferred events 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 EoD adaptation greatly improves users' satisfaction on semantic quality even if the bandwidth declines to 80kbps.

7. CONCLUSIONS

A robust Event on Demand adaptation system is achieved to support users preference when accessing video under time-variety network condition. Our proposed Event on Demand video adaptation system follows MPEG-21 digital item adaptation framework, which provides a generic adaptation solution to take account of users' preference, semantic aspects, and network environments etc. Event selection and frame dropping are effective and efficient ways to meet users' preference and adapt to the variation of network condition. MPEG-21 digital item adaptation helps to reduce computational complexity through XML manipulations. Moreover, XML based adaptation provides a generic solution for all video formats. Compared to other adaptation methods, XML based adaptation provides a quick, affordable and convenient solution.

Future work will be on designing more intelligent AQoS. More information will be considered for adaptation, such as,

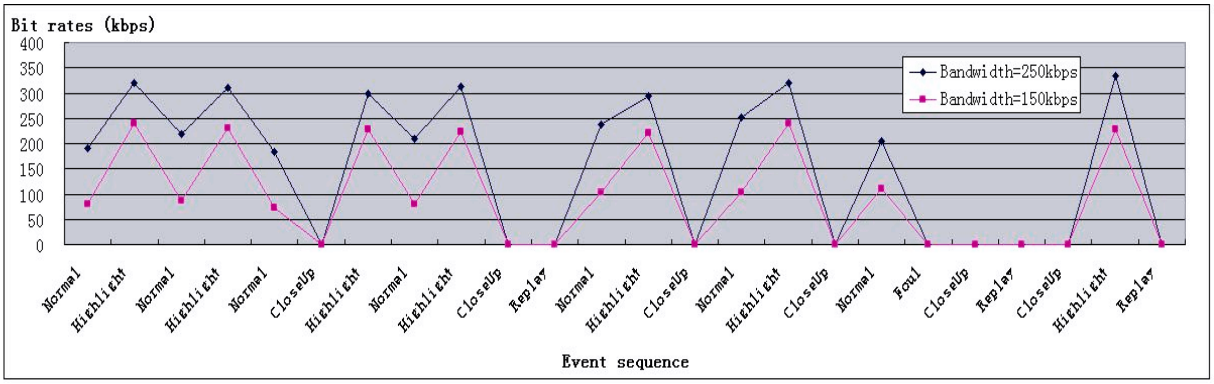

Figure 13: Curves of adapted video quality

frame size, spatial quality, etc. MEPG-21 based adaptation provides a generic solution for various resources, though we have investigated only video adaptation. In our future work, other multimedia modalities adaptation will be implemented to fulfill cross-media adaptation.

8. REFERENCES

[1] 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.

[2] Information technology - multimedia framework (mpeg-21)-part 8: Reference software. *ISO/IEC Final Standard Draft ISO/IEC 21000-8:2005(E), ISO/IEC JTC 1/ SC 29/WG 11/N5895*, 2005.

[3] E. Amir and S. McCanne. An active service framework and its application to real-time multimedia transcoding. *SIGCOMM*, 28(4):178–189, October 1998.

[4] S. Benyaminovich, O. Hadar, and E. Kaminsky. Optimal transrating via dct coefficients modification and dropping. *Proc. of 3rd Conference on Information Technology: Research and Education*, pages 100–104, June 2005.

[5] S.-F. Chang, D. Zhong, and R. Kumar. Real-time content-based adaptive streaming of sports video. *IEEE Workshop Content-Based Access to Video/Image Library,IEEE CVPR Conf.*, December 2001.

[6] D. Cotroneo, G. Paolillo, C. Pirro, and S. Russo. A user-driven adaptation strategy for mobile video streaming applications. In *Proceedings of 25th IEEE International Conference on Distributed Computing Systems Workshops*, pages 338–344.

[7] K.-T. Fung, Y.-L. Chan, and W.-C. Siu. New architecture for dynamic frame-skipping transcoder. *IEEE Trans. on Image Processing*, 11(8).

[8] J. Gecsei. Adaptation in distributed multimedia systems. *IEEE Multimedia*, 4(2):58–66, April 1997.

[9] V. K. Goyal. Multiple description coding: Compression meets the network. *IEEE Signal Processing Magazine*, 18(5):74–93, September 2001.

[10] R. Han, P. Bhagwat, R. Lamaire, T. Mummert, V. Perret, and J. Rubas. Dynamic adaptation in an image transcoding proxy for mobile web browsing. *IEEE Personal Communications*, 5(6):9–17, December 1998.

[11] M. Hicks, A. Nagarajan, and R. van Renesse. User-speciified adaptive scheduling in a streaming media network. In *Proceedings 2003 International Conference on Multimedia and Expo*, pages 281–284.

[12] S. Khan, K. F. Li, and E. G. Manning. The utility model for adaptive multimedia systems. In *International Conference on Multimedia Modeling*, pages 111–126.

[13] J.-G. Kim, Y. Wang, and S.-F. Chang. Content-adaptive utility-based video adaptation. In *Proceedings 2003 International Conference on Multimedia and Expo*, pages 281–284.

[14] W. Li. Overview of fine granularity scalability in mpeg-4 video standard. *IEEE trans. on Circuits and Systems for Video Technology*, 11(3):301–317, March 2001.

[15] S. Liu, M. Xu, H. Yi, L.-T. Chia, and D. Rajan. Multi-modal semantic analysis and annotation for basketball video. *Special Issue on Information Mining from Multimedia Databases of EURASIP Journal on Applied Signal Processing*.

[16] B. S. Manjunath, P. Salembier, and T. Sikora. Introduction to mpeg-7: Multimedia content description interface. 2002.

[17] M. Metso, A. Koivisto, and J. Sauvola. A content model for the mobile adaptation of multimedia information. *Journal of VLSI Signal Processing*, 29(1):115–128, August 2001.

[18] D. Mukherjee, G. Kuo, and A. Said. Structured scalable meta-formats (ssm) version 2.0 for content agnostic digital item adaptation - principles and complete syntax. April 2003.

[19] M. Naghshineh and M. Willebeek-LeMair. End-to-end qos provisioning in multimedia wireless/mobile networks using an adaptive framework. *IEEE Communication Magazine*, 35(11):72–81, November 1997.

[20] G. Panis and A. H. et.al. Bitstream syntax description: A tool for multimedia resource adaptation within mpeg-21. *Signal Processing: Image Communication*, 18(8):721–747, September 2003.

[21] J.-T. Park, J.-W. Baek, and J. W.-K. Hong. Management of service level agreements for multimedia internet service using a utility model. *IEEE Communications Magazine*, 39(5):100–106, May 2001.

[22] K. Shen and E. J. Delp. Wavelet based rate scalable video compression. *IEEE trans. on Circuits and Systems for Video Technology*, 9(1):109–122, Febrary 1999.

[23] X. Wang and H. Schulzrinne. An integrated resource negotiation, pricing, and qos adaptation framework for multimedia applications. *IEEE Journal on Selected Areas in Communications*, 18(12):2514–2529, December 2000.

[24] Z. Wang and C. Jan. Quality of service routing for supporting multimedia applications. *IEEE Journal on Selected Area Communications*, 14(7):1228–1234, September 1996.

[25] J. Xin, C. W. Lin, and M.-T. Sun. Digital video transcoding. *Proc. of the IEEE*, 93(1):84–97, Janaury 2005.

[26] W. Yuan, K. Nahrstedta, S. V. Advea, D. L. Jonesb, and R. H. Kravetsa. Design and evaluation of a cross-layer adaptation framework for mobile multimedia systems. In *Proceeding of SPIE-the International Society for Optical Engineering*, pages 1–13.