An Event-driven Sports Video Adaptation for the MPEG-21 DIA Framework

Min Xu, Jiaming Li, Yiqun Hu, Liang-Tien Chia, Bu-Sung Lee, Deepu Rajan, Jianfei Cai
Center for Multimedia and Network Technology, School of Computer Engineering,
Nanyang Technological University, Singapore
Email: {mxu, liji0006, y030070, asltchia, ebslee, asdrajan, asjfcai}@ntu.edu.sg
Phone: +65 67906579

Abstract

We present an event-driven video adaptation system in this paper. Events are detected by audio/video analysis and annotated by the description schemes (DSs) provided by MPEG-7 Multimedia Description Schemes (MDSs). And then, adaptation take account of users' preference of events and network characteristic to adapt video by event selection and frame dropping as following 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. Different from most existing adaptation work, the system adapts video by interesting events according to users' preference. To achieve a generic adaptation solution, the system is developed following MPEG-7 and MPEG-21 standards. gBSD based adaptation avoids complex video computation. 30 students from various departments test the system with satisfaction. Although, the system is tested on basketball video adaptation so far, it is easy to extend to other video domains.

1. Introduction

With the increasing amount of multimedia data and the development of multimedia communication techniques, developing effective and efficient video adaptation systems for users to access multimedia data according to their preference is an active research area.

Video adaptation is still a challenging field, earlier work tried to transcode video from one format to another in order to make the video compatible with the new usage environment [1]. A popular adaptation approach is to select, reduce or replace some video elements, such as dropping shots and frames in a video clip [2], drop pixels and coefficients in an image frame[3], replacing video sequence with still frames [4] etc. Although these methods provide feasible ways for video adaptation, there are still some limitations as follows: 1) most existing adaptation systems currently focused on achieving a certain defined SNR or bitrate and thus loss sight of users preference and users experience; 2) the current media adaptation solutions tend to be proprietary and therefore lacks a universal framework. 3) transcoding and video elements removal will incur high computational complexity and cost.

In order to provide a generic solution to satisfy a wide variety of applications, international standards such as MPEG-7 and MPEG-21, define the technologies users need to support exchange, access, consume, trade and otherwise manipulate digital items in an efficient, transparent and interoperable way [6, 7].

Detecting semantic highlights or events in video has attracted much interest. Most of the previous methods rely on a single feature (audio, video or transcripts) and each feature will provide some hints to interesting video events or video highlights. Recently, our event detection work based on audio sounds identification and video scene detection has shown promising results [5]. According to users' preference, these detected events can be tagged with their own priorities for video adaptation, when necessary.

Our proposed event-driven video adaptation system follows MPEG-21 digital item adaptation framework, which provides a generic adaptation solution to take account of user preference, semantic aspects, and network environments etc. Using generic Bitstream Syntax Description (gBSD) which is not aware of bitstream coding format to describe structure of bitstream provides interoperability in Digital Item Adaptation (DIA). Implementing adaptation based on the gBSD instead of the video itself helps DIA engine to adapt resource 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, that enables adaptation at different levels. Compared to other adaptation methods, XML based adaptation provides a quick, affordable and convenient solution.

2. Our Adaptation System

Our proposed adaptation system has two primary processes (Fig.1): *event identification & annotation* and *MPEG-21 Digital Item Adaptation*. In this paper, we implement our adaptation system using video scenarios from basketball games.

1-4244-0367-7/06/$20.00

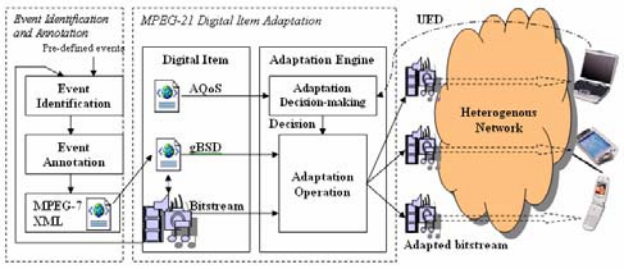

Fig. 1 Adaptation system architecture

Firstly, some pre-defined interesting events are identified by our *event identification & annotation* module and stored in MPEG-7 structured format. 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 are sent to Usage Environment Description (UED), adaptation decision engine determines decision point according to AQoS in order to maximize user satisfaction and adapt to constrained environment, such as network condition. The decision point and gBSD will instruct adaptation operation engine to alter the bitstream and resend to user.

3. Event Identification and Annotation

According to our own experience of watching videos, we actually focus on the story conveyed by the video. Event detection is significant to capture the video segments with interesting events, which can be regard as the pre-processing for video adaptation. It provides a feasible way for users to access video content by selecting their interested events.

3.1 Event Identification

Event identification is still challenging due to the gap between low-level perceptual features and high-level human perception. We are trying to seek help from some so-called middle-level features, such as some specific audio sounds and video scenes. These specific audio sounds have significant hints 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. And the excited commentator and audience sounds are most likely the aftermath 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 [5]. Four basketball events are detected as: *Replay*, *Highlight*, *Penalty* and *Normal*.

3.2 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.2.

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

Fig. 2. An example XML file of MPEG-7 event annotation

The *AudioVisual DS* is utilized to describe the temporal decomposition of a video entity. In each TemporalDecomposition DS some attributes 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, 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.

4. MPEG-21 Digital Item Adaptation

MPEG-21 digital item adaptation (DIA) framework is defined to support exchange, access, consume, trade and manipulate Digital Items in an efficient, transparent and interoperable way. We develop our system by improving Structured Scalable Meta-formats (SSM) [8] for fully content agnostic adaptation.

4.1 Generic Bitstream Syntax Description

The generic bitstream syntax description (gBSD) is an important element of Digital Item, 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 [9] 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 Section 3.2. Considering events have ranks according to various users' preference, we introduce so-call *Content-Level* to mark different events for

users accessing their events of interest. Fig.3. shows how the interaction between users defined events and metadata to 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.4.

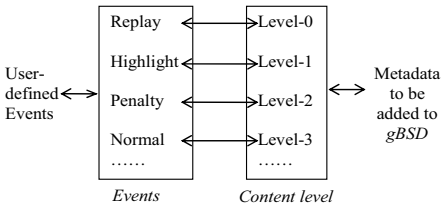

Fig. 3. The interaction between users and metadata

```
-<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" />
      …… …… ……
```

Fig. 4. An example of gBSD

4.2 Adaptation Engines

The system has two main modules which are adaptation decision and adaptation operation. The adaptation decision module is to make decision of source parameters by considering AQoS, usage environment description and constraints. Subsequently, the adaptation operation engine changes gBSD based on adaptation decision and later adapt the video resource according to the new adapted gBSD.

4.2.1 Adaptation Decision Engine

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. The adaptation decisions are made by considering three rules:

1) Event selection is based on user preference. According to users' order of preference, the parcels containing highly preferred events are most likely to remain after adaptation.
2) According to the current bandwidth, adaptation keeps as many frames as possible to convey original story.
3) With the bandwidth changing, the user's preferred segments have a higher priority of retaining all types of frames.

In the original video sequence, every gBSD unit (video segment) is associating with *Content-Level* to indicate the associated event. Once the user ranks his preferred events, the priorities of frame retaining are assigned to every gBSD unit according to user's order of preference. For example, the highest priority is assigned to those segments where the events are ranked as first order.

An example of adaptation decision making is shown in Fig.5. In this example, events' preference order selected by the user is: *Highlight*, *Normal*, *Replay* and *Penalty*. According to users' order of preference, frames in each segment will be partially or entirely dropped.

Fig. 5. Structure of original and adapted video sequence

User's order of preference is indicated in UED. AQoS associates retaining priority in the order of *Content-Level* 1, 3, 0, 2 which corresponding to different events (See Fig. 3). 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 need excessive network bandwidth. The final decision is the one utilizing most of the reserved bandwidth while preserving most of interesting events. For example, when bandwidth is 300Kbps, two possible solutions are: 1) Drop frames in the event segments with lower priority (See the first sequence in Frame dropping in Fig.5.); 2) Drop event segments with priorities lower than decision point (See the first sequence in Segment dropping in Fig.5). Of course, these two solutions can be integrated if necessary (See the second sequence in Frame dropping in Fig.5).

4.2.2 Adaptation Operation Engine

According to adaptation decisions, transformation instruction directs adaptation operation engine to alter the original gBSD and bitstream. There are two major steps:

1) *Description transformation* generates the targeted syntax description of future adapted bitstream. Transformation instruction initiates the engine to retain, delete or update gBSD units based on retaining priority in adaptation decision and annotation marking from event analysis. Frames or segments are dropped from syntax description.
2) *gBSDtoBin* parses the adapted syntax description and selects certain segments and frames to retain from the original bitstream to complete the adapted bitstream,

5. Experiments and Evaluations

An event-driven adaptation system is implemented for the MPEG-21 digital item adaptation framework. 30 students, selected from engineering and non-engineering departments, are asked to individually rank the adapted video according to their preference. We adopted the double stimulus impairment scale (DSIS) [10] method with some modifications to evaluate our adaptation system. Users view the adapted video first, followed by viewing the original video in order to avoid semantic impression from original video affecting evaluation of adapted video understanding. Two groups of experiments are conducted to evaluate the system performance.

First, we evaluate whether users are satisfied with adapted result by event dropping. 5 scales are provides for their voting. Adaptation is achieved by dropping entire event segments with lower preference during the period of limited bandwidth. The voting result (Table 1) shows that most of the students are satisfied with the event-driven adaptation as it provides them their preferred events in the limited bandwidth.

Table 1: User voting on event-driven adaptation

	Bad	Poor	Fair	Good	Excellent
Voted Result	0.0%	10.0%	23.3%	36.7%	30.0%

Second, 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 semantic understanding evaluation 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. The original video stream needs 10 seconds to transmit via 500Kbps network. We introduce 2 adapted versions with 300Kbps and 150Kbps for transmitting in 10 seconds. Table 2 shows the voting result of the 2 adapted video streams.

Table 2: User voting on frame dropping adaptation

	Bad	Poor	Fair	Good	Excellent
Bandwidth=300Kbps	0.0%	6.7%	16.7%	46.6%	30.0%
Bandwidth=150Kbps	3.3%	13.3%	30.0%	36.7%	16.7%

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 (ie. semantic quality) of the adapted video

6. Conclusions and Future Works

Event selection and frame dropping are the 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. So far, the proposed system works well for basketball video adaptation. It is easy to be extended to other video domains because of the generic solution based on MPEG-21 framework.

MEPG-21 based adaptation provides generic solution for various resources, but we have investigated only on video adaptation. In our future work, other multimedia modalities adaptation will be implemented to fulfill cross-media adaptation.

7. REFERENCES

[1] J. xin, C. –W. Lin and M.-T. Sun, "Digital video transcoding," in Proc. of the IEEE, Vol 93, no.1, pp.84-97, Jan, 2005.

[2] K.-T. Fung, Y.-L. Chan and W.-C. Siu, "New architecture for dynamic frame-skipping transcoder,", IEEE Transactions on Image Processing, Vol.11, No.8, August 2002.

[3] S. Benyaminovich, O. Hadar, E. Kaminsky, "Optimal transrating via DCT coefficients modification and dropping," in Proc of 3rd conference on Information Technology: research and education, pp:100-104, June 2005.

[4] 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., Honolulu, Hawaii, Dec. 2001.

[5] S. Liu, M. Xu, H. Yi, L.-T. Chia and D. Rajan, "Multi-modal Semantic Analysis and Annotation for Basketball Video," in Special Issue on Information Mining from Multimedia Databases of EURASIP Journal on Applied Signal Processing.

[6] B. S. Manjunath, P. Salembier, and T. Sikora, Introduction to MPEG-7, 2002.

[7] 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

[8] 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,"

[9] G. Panis, A. Hutter et.al, "Bitstream syntax description: a tool for multimedia resource adaptation within MPEG-21", 2003

[10] Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-10, ITU Telecom. Standardization Sector of ITU, August 2000.