Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Image and Signal Processing (CIISP 2007)

Video Adaptation based on Affective Content with MPEG-21 DIA Framework

Min Xu, Suhuai Luo and Jesse Jin
School of Design, Communication and Information Technology
University of Newcastle, Callaghan NSW 2308, AUSTRALIA
Email: M.Xu@studentmail.newcastle.edu.au; Suhuai.Luo, Jesse.Jin@newcastle.edu.au
Tel:61-2-49854509
Fax:61-2-49215896

Abstract

We present a video adaptation system which takes account of users' preference on video Affective Content (AC) and limited network resource. AC directly causes an user's attention, evaluation and memory which also provides feasible entry for video highlight. According to user's preference, the proposed adaptation insures the video parcels with AC are allocated as much as possible network resource. The system is implemented with MPEG-21 Digital Item Adaptation (DIA) framework which provides a generic video adaptation solution for all video formats and various usage environments by manipulating on XML files. XML file based adaptation avoids complex video computation. 30 students from various departments were invited to test the system and their responses were positive.

1 INTRODUCTION

With the increasing amount of multimedia data and the development of multimedia communication techniques,more and more users access and interact with multimedia content on different types of terminals and networks. Therefore, there is an increasing need to develop effective and efficient video adaptation systems.

Most of the adaptation work focus on developing techniques and platforms to make adaptation possible. Earlier work developed schemes of encoding to reduce video size or provide the scalability for video adaptation [6] or transcoding to make the video compatible with the new usage environment [9]. Recently, a popular adaptation approach is to select, reduce or replace some video elements, such as dropping shots and frames in a video clip [5], dropping pixels and DCT coefficients in an image frame[3], replacing video sequences with still frames [4] etc. Although these work made great contributions to achieving adaptation, there exist two limitations: 1)They focus on achieving a certain defined SNR or bitrate, lacking consideration of video content and users' preference or experience. 2) Most of the work depends on the video coding format. It lacks a generic adaptation solution.

The video adaptation is essentially a kind of interaction between human and video data under the limited resources, i.e. network conditions and terminal capabilities. In this paper, we introduce Affective Content (AC) which may cause audiences' strong reactions or special emotional experiences, such as excited sports segments, laughable segments etc. to bridge the gap between user's experience and video content. The AC not only directly affects users' experience but also provides a criterion to classify video into different genres such as horror movie, comedy and so on. Moreover, video highlight always locates in the video segments with AC. For example, sports highlight locates in excited segments while horror segments are always the highlights of horror movies. Therefore, the adaptation system conveys the video content as much as possible by paying more attention to AC and insuring the high information remaining for the video segments with AC.

To provide a generic solution, the proposed video adaptation is implemented with MPEG-21 Digital Item Adaptation (DIA) framework. Different from traditional adaptation methods, MPEG-21 DIA achieves adaptation through the manipulation on generic Bitstream Syntax Description (gBSD), which is not aware of bitstream coding format. Moreover, implementing adaptation based on gBSD instead of the video itself helps in adapting resources quickly with minimal computation cost since it alleviates the computation complexity of treating bitstream in a bit-by-bit manner.

2 OUR ADAPTATION SYSTEM

An overview of the whole system architecture is shown in Fig. 1. The resources are composed of videos which are stored in database sorted by video genres. In this paper, we use three genres of sports, horror movies and comedy

1-4244-0707-9/07/$25.00 ©2007 IEEE

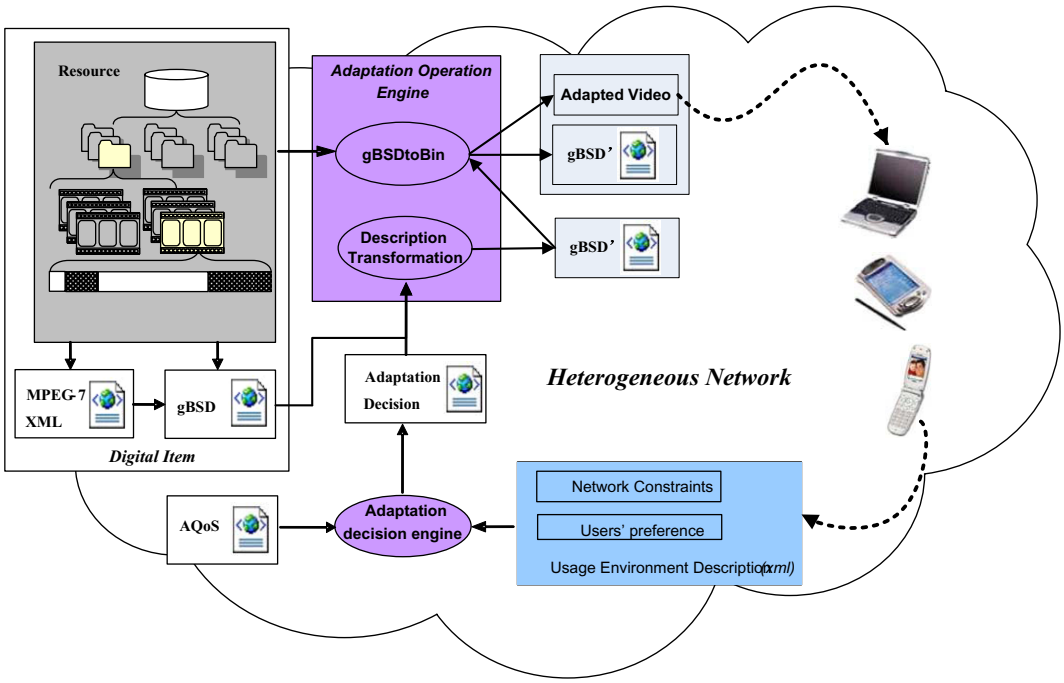

Figure 1. Adaptation system architecture

to demonstrate the proposed system. The work can be easily extended to other video domains. The AC related information such as location, duration and so on are identified by AC Identification and later stored in MPEG-7 XML files. Later on, gBSD generation module parses the AC related information from MPEG-7 XML files together with resource bitstream information to generate gBSD. According to users' preference on affective content, the video segments are tagged with different priorities for video adaptation. When segments priority, device capabilities and network conditions are sent to Usage Environment Description (UED), adaptation decision engine determines decision point according to Adaptation QoS (AQoS) in order to maximize users' satisfaction and adapt the constrained environment. Finally, the decision point and gBSD will instruct the adaptation operation engine to alter the bitstream and resend to user. The proposed adaptation is achieved by using [7], with the significant improvements as follows.

1) As a pre-processing step, Affective Content Identification and Annotation module highlights the segments with affective content, which makes users directly accessing preferred video content possible.

2)The gBSD structure and descriptions designed for easy storing and parsing of both bitstream format related information and affective content related information.

3) According to users' preference, AQoS is designed to allocate limited network bandwidth to affective content as much as possible.

4) The current network condition is estimated by monitoring the transmission time and file size of past video segment. 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.

3 DIGITAL ITEM (DI) PREPARATION

A DIGITAL ITEM is a structured digital object with a standard representation, identification and meta-data within the ISO/IEC 21000 framework [1]. This entity is also the fundamental unit of distribution and transaction within this framework. In our system, the DI includes resource, resource annotation files and resource gBSDs. In this section, we will introduce some details of resource annotation and gBSD generation.

3.1 Affective Content Identification and Annotation

Like semantic analysis, the affective content identification is challenging due to the gap between low-level perceptual features and high-level human perception. Limited video affective analysis work focuses on visual feature. In our previous work, audio and script analysis is proved to be effective to identify affective content [10, 11]. In this application, the AC Identification preforms to identify excited

segments in sports video, horror segments in horror movie and laughable segments in comedy.

To make the video structure information and AC identification results be clearly annotated and easily parsed, we store the related information in MPEG-7 XML files. MPEG-7 is 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 location and duration of AC. In detail, 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, videos with affective content are represented in a standardized and highly structured format, witch can be easily and directly parsed to gBSD generation.

3.2 Generic Bitstream Syntax Description

To generate an appropriate video description which contains the information of video format structure and affective content's location and duration is an important and necessary step for further adaptation. 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 [8] are used for parsing a bitstream to generate its gBSD description.

The bitstream is described based on parcels. In our implementation, each parcel corresponds to a video segment with or without affective content. Considering video segments have priorities according to various users' preference on AC, we introduce so-called Affective -Level to mark whether each parcel is AC or not. Affective-0 is used to indicate Non-AC while Affective-1 indicates AC.

Furthermore, frame dropping is a feasible way to adapt 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. 2.

```
... ... ... ...
<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="Affective-1">
    <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 2. An example of gBSD

4 MPEG-21 DIGITAL ITEM ADAPTATION (DIA)

MPEG-21 DIA specifies the syntax and semantics of tools that may be used to assist the adaptation of DI. The main task of MPEG-21 DIA is actually generating adapted video by selecting video elements in each parcel to meet varying usage environment conditions and maximally satisfy users' preference. Different from other adaptation methods, XML files play 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, the information which affects adaptation is conveyed between adaptation engine and the media server or media receiver instead of processing the video itself.

4.1 Usage Environment Description

In our case, Usage Environment Description (UED) refers to metadata that specifies user characteristics, terminal capabilities and network characteristics which are the limited environment constraints for adaptation decision.

Users' characteristics specify general user information, users' preference and usage history etc. In the proposed system, users are required to input their preferred video genres, video name and indicate their preference on affective content by selecting one of the following three options: View All, Highlight AC and Only AC. The user's inputs provide significant cues to set video parcel priority of allocating limited resources. (See Table 1)

Network is the medium for multimedia transmission. Heterogeneous network structure requires the transmission of multimedia files to adapt 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 described by network minimum and maximum capability and the average available bandwidth. Since the device may be provided by various network, such as LAN, MANET, Internet, wireless network, cellular network, etc, the initial network condition values are set by the profile of current access network. Whereafter, the values are updated with the degraded 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.

A usage environment description (UED) XML is generated to store usage related parameters including users preference and network conditions.

4.2 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 which are user's preference on AC and network condition. Dropping frame or whole segments is the feasible adaptation operation.

After user selects preferred video germs and video name, a certain video clip waits for adaptation. To set priorities to each parcel according to users' preference, we set 5 scales from $0 \sim 4$ as shown in Table 1. The higher priority will possibly remain more frames when frame dropping.

Table 1. Video Segments' Priority Setting Criterion

User's input	View All		Highlight AC		Only AC	
	AC	Non	AC	Non	AC	Non
Priority	2	2	3	1	4	0

On the other hand, we divide the network bandwidth into 5 scales: 0 (below 50Kbps); 1 ($50Kbps \sim 100Kbps$); 2 ($100Kbps \sim 200Kbps$); 3 ($200Kbps \sim 300Kbps$); 4 (above 300Kbps). In our case, adaptation operation is to drop different portion of videos. The 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).

The AQoS is defined by considering three rules:

1) The parcels containing the video segments with high priority 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 video segments with AC have higher priorities of retaining all types of frames.

Table 2 shows an example of feasible AQoS. According to users preference and the current network condition, decision engine finds the optimal adaptation scheme from AQoS to not only satisfy all constraints but also maximize or minimize optimization value.

Table 2. An example of feasible AQoS

Frame Dropping Scales		Network Condition Scales				
		0	1	2	3	4
Event Priority Scales	0	0	0	0	1	1
	1	0	0	1	1	2
	2	0	1	1	2	2
	3	1	1	2	2	3
	4	2	2	2	3	3

4.3 Adaptation Operation

Adaptation operation is conducted based on adaptation decision. Adaptation operation engine alters the original gBSD and bitstream by two major steps:

- *Description transformation*: Transformation instruction initiates the engine to retain, delete or update gBSD units according to adaptation 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.
- *gBSDtoBin*: This part generates the final adapted video based on the adapted gBSD structure from Description transformation. It parses the description of target adapted file to understand the structure of it. Based on the altered structure of adapted gBSD, the gBSDtoBin selects, drops or changes certain frames in bitstream.

To achieve 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, adaptation operation engine re-parses the latest adaptation decision file and performs adaptation operation based on the new adaptation rules.

5 SYSTEM EVALUATION AND CONCLUSIONS

The traditional adaptation methods which ignore users' preference on video content are evaluated by PSNR. Un-

fortunately, PSNR may not be able to achieve a reasonable evaluation for the proposed adaptation. The users' preference is an important role on the proposed adaptation system, which is a subject concept depending on individual's understanding and perception. In this case, we borrow the double stimulus impairment scale (DSIS) [2] to carry a user study in 30 students who are selected from both engineering and non-engineering departments. Each user compares adapted video with original video and vote their satisfaction for the adapted video clip based on the 5 scales from "Bad" to "Excellent". Through the user study we evaluate and compare users' satisfaction with the following two cases: adaptation only to satisfy the variation in network conditions and adaptation considering both the AC and network conditions. All the students prefer the adaptation with considering affective content since by this way, more network resource are allocated to the AC which are the segments attracting more user's attention. Table 3 shows the voting result of adaptation considering AC.

Table 3. Satisfaction Voting on AC-based 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%

Obviously, network degradation affects the user's satisfaction of the adapted video. However, the high priority assigned to retaining more frames in AC 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 of the adapted video.

6 CONCLUSIONS

A robust affective content based video adaptation system is achieved with MPEG-21 DIA framework. The main contributions of this work are: 1) AC is introduced to bridge the gap between users' experience and video content and accordingly realize the interaction between user and multimedia data. 2) Our adaptation is implemented with MPEG-21 DIA frameword, which provides a generic adaptation solution to various media formats and various usage environments. 3) The proposed adaptation provides a quick, affordable and convenient solution which helps to reduce computational complexity through XML manipulations.

Future work will be on designing more intelligent AQoS. More information will be considered for adaptation, such as, frame size, spatial quality, etc. MEPG-21 based adaptation provides a generic solution for various media resources, though we have investigated only video adaptation. Other multimedia modalities adaptation will be implemented to fulfill cross-media adaptation. The work will be extended to more video domains.

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] 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.

[3] 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.

[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.*, December 2001.

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

[6] 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.

[7] 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.

[8] 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.

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

[10] M. Xu, L.-T. Chia, and J. Jin. Affective content analysis in comedy and horror videos by audio emotional event detection. In *Proceedings of IEEE International Conference on Multimedia and Expo*.

[11] M. Xu, L.-T. Chia, H. Yi, and D. Rajan. Affective content detection in sitcom using subtitle and audio. In *Proceedings of the 12th IEEE International Multimedia Modelling Conference*.