계층화 비디오 브로드캐스팅을 위한 QoS 적응변환방법

장콩탕 강정원 이경준 유정주 임종수
한국전자통신연구원
{tcthang, jungwon, lkj0610, jjyoo, ljten}@etri.re.kr

Scalable Video Broadcasting with QoS Adaptation

Truong Cong Thang Jung Won Kang Kyung Jun Lee Jeong Ju Yoo Jong Soo Lim
Electronics and Telecommunications Research Institute

요약

Modern broadcasting/multicasting networks has the heterogeneous nature in terms of terminals and available bandwidth. Such heterogeneity could be coped by scalable video coding (SVC) standard developed recently. More specifically, spatial layers of an SVC bitstream can be consumed by different terminals and SNR (and temporal) scalability can be used to cope with bandwidth heterogeneity. In this work, we tackle the problem of SVC adaptation for different user groups receiving the same broadcast/multicast video, so as to provide a flexible tradeoff between the groups while also maximizing the overall quality of the users. The adaptation process to truncate an SVC bitstream is first formulated as an optimization problem. Then the problem is represented by MPEG-21 DIA description tools, which can be solved by a universal processing. The results show that MPEG-21 DIA is useful to enable automatic and interoperable adaptation in our scenario.

1. Introduction

In modern broadcasting and multicasting networks (e.g. Digital Multimedia Broadcasting or IPTV), terminals of end-user could have much different capabilities and characteristics. For example, to receive the same video channel, one user group may use PMPs while another group uses mobile phones; also the number of users in one group may widely vary according to time and location. Moreover, in multicasting and narrowcasting cases, the bandwidth available for a video channel could be dynamically changed due to actual conditions of networks. Such heterogeneity could be coped by using scalable coding formats, especially the scalable video coding (SVC) standard recently developed by the joint video team (JVT) of MPEG and ITU-T [1].

The scalability of SVC is possible in 3 dimensions: spatial, temporal, and SNR. Though SVC provides simple and flexible truncation of a coded bitstream, there still need to be various tools and methods to support the adaptation in different scenarios. Currently an adapted SVC bitstream is mostly targeted at a single kind of terminal. Whereas, it is obvious that spatial layers of an SVC bitstream can be consumed by different terminals and SNR (and temporal) scalability can be used to cope with bandwidth heterogeneity.

In this work, we tackle the problem of SVC adaptation for different user/terminal groups receiving the same broadcast or multicast video. Our goal is to provide a flexible tradeoff between the user groups while also maximizing the overall quality of the users. As an example scenario, suppose that a broadcast or multicast SVC video is encoded with two spatial layers (QCIF&CIF), both enhanced by SNR enhancement (SE) data. One user group uses mobile phone to decode the QCIF layer and the other group uses PMP to decode the CIF layer. The SE data may be truncated to meet a bitrate (bandwidth) constraint allocated to that "video channel".

To provide a tradeoff between the user groups, we present a systematic approach based on MPEG-21 DIA [2] to truncate SE data at different spatial layers of an SVC bitstream. The adaptation process is first formulated as an optimization problem. Then this problem is represented by MPEG-21 DIA description tools, which can be solved by a universal processing. The results show that MPEG-21 DIA is useful to enable automatic and interoperable adaptation in our scenario.

2. Framework Description

2.1 Overview

The overall diagram of our system is shown in Fig. 1. The server includes three main modules, namely adaptation decision taking engine (ADTE), extractor, and streamer. At the client side, there are some user groups using different terminals.

Each input SVC bitstream is augmented with an AdaptationQoS description. This description shows the adaptation behavior, specifically the relation of adaptation choices (operators), associated quality, and resource requirement of the bitstream. The "structured wrapper" of the bitstream and associated descriptions is called "Digital Item" (DI) in MPEG-21.

The usage environment descriptions (UED), which are provided by the streamer, describe the characteristics of networks, terminals, and users. These descriptions are crucial in recognizing the constraints of adaptation. The system also store one or more universal constraint descriptions (UCD) which describes the adaptation goals and limits for each session.

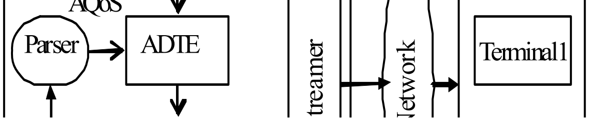


Fig. 1. Framework of MPEG-21 enabled SVC adaptation

The ADTE takes as input the AdaptationQoS, UCD, and UED descriptions. Parser modules are used to parse the XML (extended mark-up language) messages containing description. Based on these descriptions, ADTE makes decisions and provides as output the values of operators. Then the extractor truncates the bitstream according to the decided values of operators. Finally, the streamer module sends the adapted bitstream to user groups. We assume that each user group views one spatial layer of the video.

2.2 Problem Formulation

Denote R_o and R^c the original bitrate and the bitrate constraint of the bitstream to be adapted. Also denote N the number of spatial layers of the bitstream. For spatial layer i (i=1,..,N), denote p_i the adaptation operation applied to that layer, which results in quality Q_i and bitrate R_i. The overall quality of a bitstream is denoted as OQ. The problem formulation is defined as follows:

Find the optimal operations $\{p_i\}$ that:

$$\text{maximize} \quad OQ \qquad (1)$$

while satisfy $\sum_{i=1}^{N} R_i \le R^c$. (2)

We currently define the overall quality as:

$$OQ = \sum_{i=1}^{N} w_i \cdot Q_i \qquad (3)$$

where w_i is the weight of layer i ($0 \le w_i \le 1$).

In general, R_i and Q_i are functions of $\{p_i\}$. Currently, we let p_i be the *truncated bitrate* of SNR enhancement data of spatial layer i. So, Eq. (2) can be rewritten as:

$$R_o - \sum_{i=1}^{N} p_i \le R^c \quad . \qquad (4)$$

In SVC, higher spatial layer signal may be predicted from lower spatial layers. Due to this fact, Q_i will depend on not only p_i of current spatial layer but also p_{i-1}, p_{i-2},…, p_1 of lower spatial layers, That means,

$$Q_i = g(p_i, p_{i-1}, \ldots, p_1). \qquad (5)$$

Eq. (5) actually represents the R-D information of a spatial layer. In practice, this information is usually in discrete form.

Note that with the above scenario, the extreme cases of weight values, i.e. $\{w_1=1, w_2=0\}$ or $\{w_1=0, w_2=1\}$, correspond to QCIF-max and CIF-max truncation methods [3]. Obviously, these extreme methods try to maximize the quality of only one spatial layer (either QCIF or CIF), regardless of the other layer.

Usually, if one user group is more important than other groups, that group should be provided with a better quality. It is expected that by adjusting the values of w_i's in this formulation, we can flexibly and optimally provide a tradeoff between spatial layers when they are consumed by different user groups.

Other optimization problems in practice may have different parameters and complexity; however, they in general still need some kinds of R-D information, several optimization criteria and limit constraints. In MPEG-21 DIA [2], a variety of description tools (i.e. metadata) has been developed to support adaptation systems where all these factors are represented by standardized metadata, thus enabling the interoperability of future multimedia communication.

3. System Implementation

This section focuses on the specific techniques used to achieve the adaptation goal described above. Detailed information about the architecture of our system can be found in [4][6].

3.1 AdaptationQoS Description

As mentioned, an AdaptationQoS description describes the relationships between the possible adaptation operator values, the associated quality values and resource requirements. Because these relationships are not easy to be obtained in real-time, such kind of metadata could be the only means to support online adaptation in practice.

An AdaptationQoS description may consist of a number of modules, each can take one of three formats: utility function, look-up table, or stack function. Utility function describes a list of adaptation points and look-up table is a matrix representation. Meanwhile, stack function allows data representation in the form of parametric equations. More details of MPEG-21 DIA in general and AdatpationQoS tool in particular can be found in [5][2].

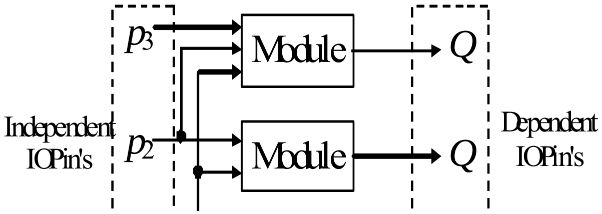


Fig. 2. Composition of AdaptationQoS description for SVC video

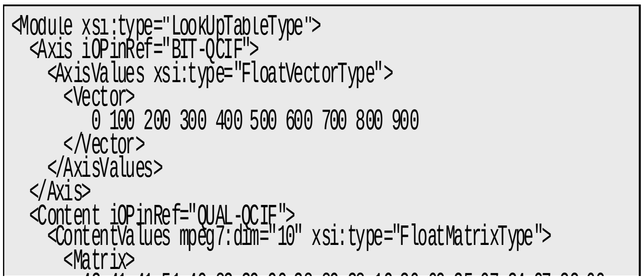
Given an original bitstream in our scenario, the operational R-D data (i.e. Eq. (5)) of the spatial layers can be computed in advance and stored using AdaptationQoS tool of MPEG-21 DIA. The overall mechanism is shown in Fig. 2 for the case of a bitstream with 3 spatial layers. Here, each spatial layer is represented by an *AdaptationQoS module*. Modules 1, 2, 3 respectively correspond to the first, the second, and the third (or highest) spatial layers. For each module, its inputs (independent IOPins) are the operations of the corresponding layer and its lower layers, while the output (dependent IOPin) is the adapted quality.

Fig. 3 and Fig. 4 show examples of AdaptationQoS modules of a bitstream consisting of CIF layer and QCIF layer. Essentially, these two modules respectively show the relationships $Q_2 = g(p_2, p_1)$ and $Q_1 = g(p_1)$. These modules are actually the description of Harbour sequence which will be used in the experiment section.

In Fig. 3, the IOPin's `BIT-QCIF, BIT-CIF, QUAL-CIF`, correspond to p_1, p_2, and Q_2 of CIF layer. The values of p_1 and p_2 take unit of Kbps and actually have a step size of 100Kbps. The 10x10 matrix of this module represents the values of Q_2 for each (p_2, p_1) pair. The unit of Q_2 and Q_1 is average PSNR. Similarly, in Fig. 4, the AdaptationQoS module of QCIF layer has two IOPin's, `BIT-QCIF` and `QUAL-QCIF`, corresponding to p_1 and Q_1.

```
<Module xsi:type="LookUpTableType">
  <Axis iOPinRef="BIT-CIF">
    <AxisValues xsi:type="FloatVectorType">
      <Vector>
         0 100 200 300 400 5000 600 700 800 900
      </Vector>
    </AxisValues>
  </Axis>
  <Axis iOPinRef="BIT-QCIF">
    <AxisValues xsi:type="FloatVectorType">
      <Vector>
        0 100 200 300 400 500 600 700 800 900
      </Vector>
    </AxisValues>
  </Axis>
  <Content iOPinRef="QUAL-CIF">
    <ContentValues mpeg7:dim="10 10" xsi:type="FloatMatrixType">
      <Matrix>
        33.59 33.55 33.51 33.41 33.29 33.20 32.82 32.64 32.04 31.44
        33.26 33.22 33.18 33.08 32.97 32.89 32.53 32.36 31.79 31.22
        32.98 32.95 32.91 32.81 32.70 32.63 32.28 32.12 31.59 31.04
        32.71 32.67 32.63 32.54 32.44 32.37 32.04 31.89 31.39 30.86
        32.43 32.39 32.36 32.27 32.18 32.11 31.80 31.66 31.18 30.68
        32.22 32.19 32.15 32.07 31.98 31.91 31.62 31.48 31.02 30.55
        31.93 31.90 31.87 31.80 31.71 31.65 31.37 31.24 30.80 30.35
        31.32 31.29 31.26 31.19 31.11 31.06 30.81 30.69 30.29 29.88
        30.67 30.64 30.62 30.56 30.48 30.43 30.20 30.09 29.73 29.36
        30.11 30.09 30.06 30.00 29.93 29.89 29.67 29.57 29.25 28.91
      </Matrix>
    </ContentValues>
  </Content>
</Module>
```

Fig. 3. AdaptationQoS module for CIF layer



```
<Module xsi:type="LookUpTableType">
  <Axis iOPinRef="BIT-QCIF">
    <AxisValues xsi:type="FloatVectorType">
      <Vector>
         0 100 200 300 400 500 600 700 800 900
      </Vector>
    </AxisValues>
  </Axis>
  <Content iOPinRef="QUAL-QCIF">
    <ContentValues mpeg7:dim="10" xsi:type="FloatMatrixType">
      <Matrix>
```

Fig. 4. AdaptationQoS module for QCIF layer

3.2 Constraint Composition

In MPEG-21 DIA, the constraints of an optimization problem are represented by the Universal Constraints Description tool (UCD). Constraints can be of two types. The first type is optimization constraint which aims at maximizing or minimizing a certain factor (e.g. Eq. (1)). The second type is limit constraint which is a Boolean criterion that some IOPin's should satisfy. Given the above example of 2-layer bitstream, the two constraints represented by the UCD description are:

$$Maximize \quad (w_1 \cdot Q_1 + w_2 \cdot Q_2), \quad \text{and}$$
$$(R_o - p_1 - p_2 \leq R^c) = TRUE \quad .$$

The bitrate constraint R^c is referenced from UED description provided by the streamer. The weight w_i's can be inferred from users' profiles, status, and the number of users in each group. For example, in narrowcasting, if we know who is the most important user and which terminal he currently uses, the corresponding spatial layer will be emphasized. In special cases, when a group has no users, the corresponding w_i should be set to 0. This fact implies that w_i's (and bitrate constraint R^c as well) may *vary both among sessions and during a session*, and thus the ADTE should quickly respond to any of these changes.

3.3 Optimization Strategy

For seamless adaptation, it is important that the processing time of ADTE should be small. The advantage of MPEG-21 enabled approach is that the metadata-represented problem actually can be solved by a universal decision-making process with different optimization strategies as generally sketched in [5]. In our problem formulation, the complexity of solution searching depends on the number of spatial layers N and number of operation choices (denoted as $C(p_i)$) in each spatial layer. For a practical SVC bitstream, the highest value of N is only 3. Meanwhile, from our experience, $C(p_i)$ is less than several dozens as the human often cannot differentiate many visually-similar adapted versions. To speed up the exhaustive search, we have employed the Viterbi algorithm of dynamic programming [3] in our system. With the settings N=3 and $C(p_i)$=100, the processing time of ADTE is found to be below 15ms. This value is negligible for both one way and two-way communications.

3.3 Dynamic Extractor

Based on the description tools, the ADTE determines the optimal amount of truncated bitrate $\{p_i\}$. This information is then pushed to the extractor to carry the adaptation at bitstream level. The dynamic extractor in our previous work [6] is extended a little to truncate SNR enhancement (SE) data at multiple spatial layers. Now, the truncation ratio of SE NAL units of spatial layer i is computed using the corresponding p_i. As the modification is small, the extractor still works in a dynamic and real-time manner [6].

4. Experiments

This section provides some experiment results that show the effectiveness of the proposed method. Test video sequences are encoded by JSVM7.12 with 2 spatial layers, QCIF and CIF. SNR enhancement for each spatial layer is FGS mode. The quantization parameter of each base quality layer is 38. The size of GOP is 16. Quality metric used in the experiments is average PSNR. Suppose that two user groups will consume an adapted bitstream as in the above scenario.

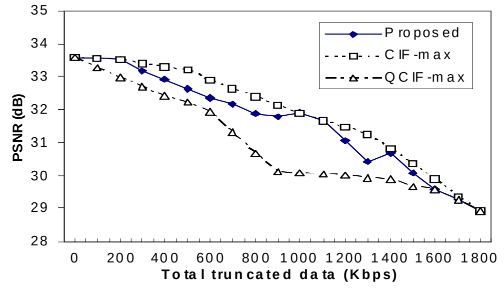
The AdaptationQoS modules in Fig. 3 and Fig. 4 are the R-D data of the Harbour test sequence, with truncation step of 100Kbps. Specifically, each p_1 takes value among {0, 100, 200,..., 900}. With this input bitstream, FGS bitrates of QCIF and CIF layers are 938Kbps and 906Kbps. Table I shows the operator values provided by different adaptation methods for Harbour sequence. The first column shows the total amount of truncated bitrate (in both QCIF and CIF layers). The table shows the following approaches, *CIF-max*, *QCIF-max*, and our proposed method with $\{w_1$=0.20, w_2=0.80$\}$ and $\{w_1$=0.40, w_2=0.60$\}$. As the PSNR of QCIF layer is often higher than that of CIF layer, the selected value of w_1 is often less then that of w_2 to achieve some balance between the layers.

We can see that the QCIF-max method, as expected, simply truncates up the FGS data of CIF layer first, and then QCIF layer. On the other hand, the CIF-max method always tries to truncate some FGS data of QCIF layer so as the quality of CIF layer is maximized. Meanwhile, the operator values of our proposed method provide a tradeoff between the two previous methods.

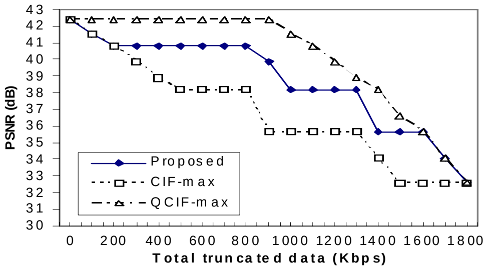
Fig. 5 compares the quality of the proposed method (when w_1=0.20, w_2=0.80) with CIF-max and QCIF-max methods. The quality curves of CIF layer (Fig. 5a) and QCIF layer (Fig. 5b) are shown with respect to the total amount of truncated FGS data in both layers. It can be seen that CIF-max and QCIF-max truncations are the two extremes cases of quality optimization; one spatial layer is maximized but the other layer is strongly degraded. Meanwhile, our proposed method provides a balance between the two extremes.

TABLE I: Operator Values Provided by Different Adaptation Methods (harbour sequence, CIF&QCIF layers)

p_1+ p_2 (Kbps)	QCIF-max (1, 0)		CIF-max (0, 1)		Proposed (0.2, 0.8)		Proposed (0.4, 0.6)	
	QCIF	CIF	QCIF	CIF	QCIF	CIF	QCIF	CIF
100	0	100	100	0	100	0	0	100
300	0	300	300	0	200	100	0	300
500	0	500	500	0	200	300	0	500
700	0	700	500	200	200	500	100	600
900	0	900	700	200	300	600	200	700
1100	200	900	700	400	500	600	200	900
1300	400	900	700	600	500	800	500	800
1500	600	900	900	600	700	800	600	900
1700	800	900	900	800	800	900	800	900



(a) Quality of CIF layer

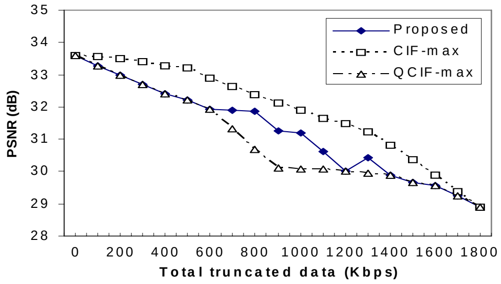


(b) Quality of QCIF layer

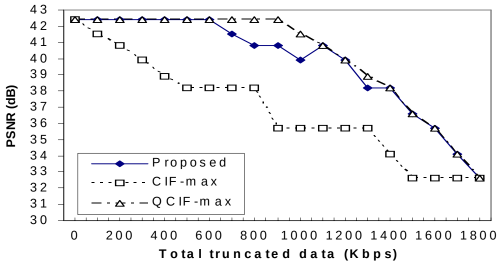
Fig. 5. Comparison of different truncation methods for Harbour sequence. The proposed method has w1=0.20, w2=0.80.

Fig. 6 compares the quality of the proposed method when w_1=0.40, w_2=0.60. Compared to previous values, these weight values emphasize somewhat the QCIF layer. As a result, in Fig. 6 the quality curves of the proposed method move closer to the QCIF-max curves. So, the tradeoff between the two extreme cases can be flexibly controlled by adjusting the weight values.

In our method the improvement of one layer leads to degradation in the other layer; however, a solution with controllable tradeoff would help avoid severe degradation of a layer and increase the overall quality of the users. It should be noted that, with the above framework, the optimality of overall quality is always guaranteed. As mentioned above, the processing time of both ADTE and extractor is small. So if there is any change in bitrate constraint and user status, the solution $\{p_i\}$ can be recomputed in real-time and the bitstream is truncated accordingly. That means the whole adaptation process is seamless to the users.



(a) Quality of CIF layer



(b) Quality of QCIF layer

Fig. 6. Comparison of different truncation methods for Harbour sequence. The proposed method has w_1=0.40, w_2=0.60.

5. Conclusions

In this paper, we have studied the adaptation of SVC bitstream for different user groups using the MPEG-21 multimedia framework. Through a controllable tradeoff between spatial layers, our proposed method helps avoid severe quality degradation for certain user group and maximize the overall quality of all users. For future work, we will consider adaptation behavior of SVC video guided by perceptual quality metrics instead of PSNR.

Acknowledgment

This work was supported by the IT R&D program of MIC/IITA. [2006-S-017-03, Development of advanced transmission technology for the terrestrial DMB system]

References

[1] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, pp. 1103-1120, Sept. 2007.

[2] ISO/IEC IS 15938-5:2001, "Information Technology – Multimedia Content Description Interface – Multimedia Description Schemes," 2003.

[3] T. C. Thang, J. W. Kang, J.-J. Yoo, Y. M. Ro, "Optimal multi-layer adaptation of SVC video over heterogeneous environments", Journal of Advances in Multimedia, 2008.

[4] J. W. Kang, S. Jung, J.-G. Kim, J. Hong, "Development of QoS-aware ubiquitous content access testbed", IEEE Trans. Consumer Electron., vol. 53, no. 1, pp. 197-203, Feb. 2008.

[5] D. Mukherjee, E. Delfosse, J.-G. Kim, and Y. Wang, "Optimal Adaptation Decision-Taking for Terminal and Network Quality-of-Service," IEEE Trans. Multimedia, vol. 7, no. 3, pp.454-462, Jun. 2005.

[6] H. Choi, J. W. Kang, D.-S. Jun, J.-G. Kim, Y. M. Ro, "Dynamic and Interoperable Adaptation of SVC for QoS-enabled Streaming", IEEE Trans. Consumer Electron., vol. 53, no. 2, pp. 384-389, May 2008.