

A Metadata-enabled Approach for Scalable Video Streaming in Heterogeneous Networks

Truong Cong Thang^{1,*}, Hung T. Le¹, Duc V. Nguyen¹, and Anh T. Pham¹

Abstract

In today's pervasive computing environments, multimedia content should be adapted to meet various conditions of network connections, terminals, and user characteristics. Scalable Video Coding (SVC) is a key solution for video communication over heterogeneous networks, where user terminals have different capabilities. This paper presents a standard-compliant approach that adapts an SVC bitstream to support multiple users. The adaptation problem is formulated as an optimization problem, focusing on the tradeoff between qualities of different spatial layers of an SVC video. Then the adaptation process is represented by standard metadata of MPEG-21, which can be solved by universal processing to enable interoperable and automatic operation. Our approach provides the users with optimal quality, a wide flexibility, and seamless adaptation. To the best of our knowledge, this is the first study that shows the adaptation tradeoff between spatial layers of a conforming SVC bitstream.

Key Words: Heterogeneous Networks, Scalable Video Coding, Video Adaptation.

I. INTRODUCTION

In pervasive computing environments, multimedia content should be adapted to meet various conditions of network connections, terminals, and user characteristics. Scalable video coding (SVC) has been shown as a good video coding format for multimedia communication over heterogeneous networks [1-2]. SVC format, which is extended from the advanced video coding (AVC) [3], is appropriate to create a wide variety of bitrates with high compression efficiency. Especially, an SVC bitstream can be easily truncated in different manners to meet various characteristics of devices and connections. The scalability is possible in 3 dimensions: temporal, SNR (or quality), and spatial. Temporal scalability is enabled by hierarchical prediction, whereas spatial scalability and SNR/quality scalability are provided for using a layered approach [3]. SNR scalability is provided in two normative modes, CGS (coarse grained scalability) and MGS (medium grained scalability), where MGS is high interest due to the advantage of packet-based scalability.

There have been a number of researches that show the usefulness of the scalability of SVC for adaptive delivery

under the constraints of bandwidth (e.g. [4-5]) and packet loss (e.g. [6]). Regarding spatial scalability of SVC, in [7], spatial resolution of video content is adapted to cope with the low bandwidth and limited power of mobile terminals. Besides, some studies employ the MPEG-21 multimedia framework to deliver media content in an interoperable manner [8][9]. In [10], the adaptation approach based on bitstream syntax description (BSD), which is a tool of MPEG-21, is applied to SVC content. Also, an extensive review of the adaptation methods for SVC can be found in [11].

Our focus in this study is on the use of spatial layers of an SVC bitstream to serve different users. As an example scenario, suppose that an SVC bitstream is encoded with two spatial layers, both enhanced by MGS data. The MGS data can be truncated to meet a bitrate constraint and then the adapted bitstream is sent to a remote place with two user groups, each receiving one spatial resolution. This case is also applicable to mobile broadcasting with different kinds of terminals [1-2]. By means of the adaptation, the quality of one layer can be sacrificed to increase the quality of the other layer, or the qualities of two layers can be kept balanced. A similar adaptation

Manuscript received March 2, 2015; Revised April 5, 2015; Accepted April 28, 2015. (ID No. JMIS-2015-0007)

Corresponding Author(*): Truong Cong Thang, Tsuruga, Ikki-Machi, Aizu-Wakamatsu, Japan 965-8580, Tel: +81-242-37-2560, E-mail: thang@u-aizu.ac.jp

¹Computer Engineering Division, The University of Aizu, Aizu-Wakamatsu, Japan, E-mail: {thang, d8162102, m5182102, pham}@u-aizu.ac.jp

problem was investigated in [12] for video streams coded with fine-grained scalability (FGS). However, it should be noted that, though FGS was standardized in MPEG-4, it was not included in the SVC specification. For multilayer SVC bitstream, the adaptation of the bitstream to support multiple users is a non-trivial task. The reason is that an SVC bitstream must strictly obey the requirement of layer dependency; any missing packets in lower layers would result in a non-compliant bitstream. To the best of our knowledge, no previous studies have shown the adaptation tradeoff between spatial layers of a conforming SVC bitstream.

In this paper, we present a systematic and standard-compliant approach to deliver an SVC bitstream to multiple users (or user groups), so as to provide a flexible tradeoff among users while maximizing the overall quality of the users. The adaptation process is firstly formulated as an optimization problem. Then this problem is represented by MPEG-21 DIA description tools, which can be solved by universal processing [13]. The results show that our system is effective to enable interoperable adaptation in heterogeneous environments.

This paper is organized as follows. Section II provides a brief overview of SVC. The description of our framework, including problem formulation, is presented in Section III. Section IV presents detailed system implementation to achieve the adaptation goals. Experiments and discussions are given in Section V. Finally, conclusions and future work are provided in Section VI.

II. SVC OVERVIEW

An SVC bitstream is divided into NAL (network abstraction layer) units. Each NAL unit containing coded video data has four key parameters for dependency information in their headers, namely `dependency_id`, `temporal_id`, `quality_id`, and `priority_id`, which indicate the identifiers of a spatial layer (or a CGS layer), a temporal layer, a quality layer, and a “priority layer”. These parameters are the basic information used to discard one or more NAL unit in the adaptation process [3].

The structure of an example SVC bitstream with MGS coding is shown in Figure 1, where each block is a NAL unit. This bitstream contains two spatial layers. Each spatial layer consists of one quality base layer and a number of MGS layers. All NAL units belonging to a picture are organized into one access unit. The dependency between access units is indicated by `temporal_id` parameter, which is assigned based on the hierarchical temporal prediction of the given bitstream.

In SVC, NAL units of an MGS layer can be individually discarded. Especially, the coded data

(including motion data and residual data) corresponding to a quantization step size can be fragmented into at most 15 MGS layers. We call the set of fragmented MGS layers corresponding to a quantization step size an “MGS stack”. Note that motion information is always stored in the first (fragmented) MGS layer. Thanks to this splitting, “finer” scalability can be achieved. This kind of packet-based scalability can be considered as a compromise between CGS and FGS. More information about SVC bitstream structure and possible adaptation options could be found in [11].

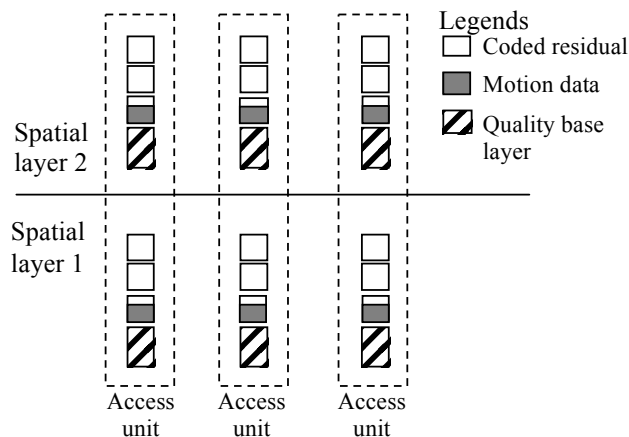


Fig. 1. Structure of an SVC bitstream with two spatial layers

As discussed in [4, 11], due to the requirement of layer dependency, SNR adaptation in SVC reference software only discards MGS data of the highest spatial layer. However, using our method proposed in [4], MGS data of lower spatial layer(s) can be discarded as well. When different spatial layers could be processed at the same time, there would be some trade-off between the layers, and thus the flexibility to adapt to specific situations. In the following, we will present a systematic approach to exploit this flexibility.

III. ADAPTATION FRAMEWORK DESCRIPTION

1. System Overview

The overall diagram of our system is shown in Fig. 2. The server includes three main modules, namely adaptation decision taking engine (ADTE), dynamic extractor, and streamer. At the client side, there are different groups of users using different terminals. Each input SVC bitstream is augmented with an AdaptationQoS description [13]. 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 [14].

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 context of adaptation. The system also stores one or more universal constraint descriptions (UCD) which describe the adaptation goals and limits for each session.

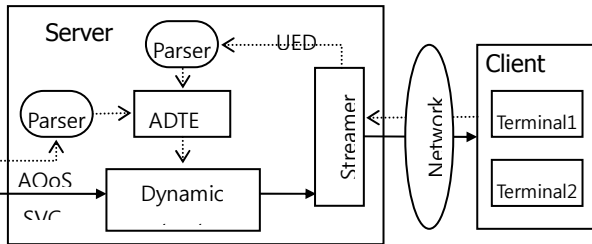


Fig. 2. Framework of MPEG-21 enabled SVC adaptation

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 the descriptions. 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 users. Here, the users use different terminals to receive the same content provided by the server. In practice, UED could be obtained from the terminals. For example, a terminal may monitor its connection bitrate and power, and then send that information in the form of UED back to the server.

2. Problem Formulation

Denote R_o and R^c the original bitrate and the constraint bitrate of the bitstream to be adapted. Also denote N the number of spatial layers of the bitstream. For spatial layer i ($i=1, \dots, N$), denote p_i the adaptation operator value 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 operator values $\{p_i\}$ that:

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

$$\text{while satisfy } \sum_{i=1}^N R_i \leq R^c \quad (2)$$

We define the overall quality as:

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

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

In general, R_i and Q_i are functions of $\{p_i\}$. Currently, we let p_i be the *discarded bitrate* of MGS data of spatial layer

i . So, Eq. (2) can be rewritten as:

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

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 $p_{i-1}, p_{i-2}, \dots, p_1$ of lower spatial layers, That means,

$$Q_i = g(p_i, p_{i-1}, \dots, p_1) \quad (5)$$

Eq. (5) actually represents the R-D information of a spatial layer. In practice, this information can be in discrete form as discussed in the next section.

Note that, with the above scenario, two extreme cases of weight values, i.e. ($w_1=1, w_i=0$ with $i=2 \sim N$) and ($w_i=0$ with $i=1 \sim N-1, w_N=1$), correspond to two special adaptation options, called *top-max* and *bottom-max*. When either of these methods is employed, only one spatial layer (namely top spatial layer with top-max option or the lowest spatial layer with bottom-max option) is maximized, regardless of the other layer.

Moreover, if one user is more important than other users, that user 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 users.

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 [14], a variety of description tools (i.e. metadata) have been developed to support adaptation systems where all these factors are represented by standardized metadata, thus enabling the interoperability of future multimedia communication.

III. SYSTEM IMPLEMENTATION

This section presents the specific techniques used to achieve the adaptation goal described above. We will focus on the use of MPEG-21 DIA descriptions, optimization strategy inside ADTE, and the dynamic extractor.

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 the three formats: utility function, look-up table, or stack function [13]. 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 AdaptationQoS tool in particular can be found in [13][14].

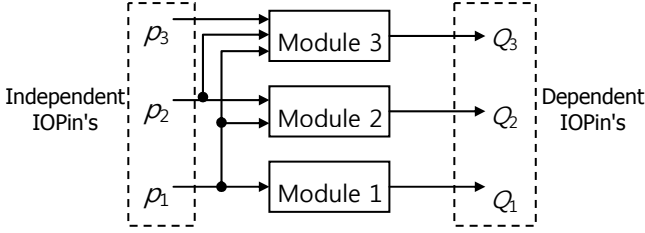


Fig. 3. Composition of AdaptationQoS description for SVC

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. 3 for the case of a bitstream having 3 spatial layers. As shown in Figure 3, each spatial layer is represented by an *AdaptationQoS module*. Here, 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 adaptation operator values of the corresponding layer and its lower layers, while the output (dependent IOPin) is the adapted quality.

As discarding of MGS data in practice is done with discrete values, the utility function or look-up table with discrete content is suitable to represent the R-D information of an SVC bitstream. Fig. 4 and Fig. 5 show examples of AdaptationQoS modules of a bitstream consisting of 2 spatial layers. 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 Football sequence which will be used in the experiment section.

In Fig. 4, the IOPin's BIT-L1, BIT-L2, QUAL-L2, correspond to p_1 , p_2 , and Q_2 of the top spatial layer. The values of p_1 and p_2 take unit of kbps and actually have a step size of 100Kbps. The 15x8 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. 5, the AdaptationQoS module of the bottom layer has two IOPin's, BIT-L1 and QUAL-L1, corresponding to p_1 and Q_1 .

Note that MPEG-21 DIA allows obtaining a finer representation by interpolating these existing points of an AdaptationQoS description. Moreover, the number of

stored points is not many as a piece-wise R-D representation could be enough in practice [15].

```
<Module xsi:type="LookUpTableType">
  <Axis iOPinRef="BIT-L2">
    <AxisValues xsi:type="FloatVectorType">
      <Vector>
        0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400
      </Vector>
    </AxisValues>
  </Axis>
  <Axis iOPinRef="BIT-L1">
    <AxisValues xsi:type="FloatVectorType">
      <Vector>
        0 100 200 300 400 5000 600 700
      </Vector>
    </AxisValues>
  </Axis>
  <Content iOPinRef="QUAL-L2">
    <ContentValues mpeg7:dim="15 8" xsi:type="FloatMatrixType">
      <Matrix>
        34.55 34.35 33.96 33.34 32.91 32.25 31.46 30.56
        34.10 33.93 33.57 33.01 32.61 32.00 31.24 30.38
        33.52 33.40 33.09 32.59 32.21 31.63 30.94 30.15
        ...
        30.23 30.20 30.10 29.92 29.77 29.55 29.25 28.88
        29.47 29.41 29.39 29.27 29.15 29.01 28.78 28.45
        28.82 28.86 28.76 28.68 28.58 28.49 28.31 28.07
        28.45 28.31 28.30 28.25 28.17 28.10 27.97 27.78
      </Matrix>
    </ContentValues>
  </Content>
</Module>
```

Fig. 4. AdaptationQoS module for top layer

```
<Module xsi:type="LookUpTableType">
  <Axis iOPinRef="BIT-L1">
    <AxisValues xsi:type="FloatVectorType">
      <Vector>
        0 100 200 300 400 500 600 700
      </Vector>
    </AxisValues>
  </Axis>
  <Content iOPinRef="QUAL-L1">
    <ContentValues mpeg7:dim="10" xsi:type="FloatMatrixType">
      <Matrix>
        35.63 34.53 33.91 32.79 32.17 31.27 30.40 29.64
      </Matrix>
    </ContentValues>
  </Content>
</Module>
```

Fig. 5. AdaptationQoS module for bottom layer

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 a 2-layer bitstream, the two constraints represented by the UCD description are:

$$\text{Maximize } (w_1 \cdot Q_1 + w_2 \cdot Q_2),$$

$$\text{and } (R_o - p_1 - p_2 \leq R^c) = \text{TRUE}.$$

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 and status. For example, if we know who the most important user is and which terminal that user currently uses, the corresponding spatial layer will be emphasized. When each spatial layer is consumed by a group of users, the weight may be inferred from the number of users in each group. This fact implies

that w_i 's (and bitrate constraint R^c as well) may vary during a session, and thus the adaptation system should quickly respond to any of these changes.

To better facilitate the processing of the system, a special UED description for the weight values can be devised. The weight values in this separate description can be computed at the provider side and/or the user side and then referenced by the UCD description in the same manner as other UCD parameters.

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 [13]. With this discrete-valued case, the simple solution of exhaustive search is expected to be suitable as the problem space in practice is often not large [13].

In our problem formulation, the complexity of solution searching depends on the number of spatial layers N and the 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 [16-17] 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. In [18], for some use cases having multiple optimization constraints, three different optimization strategies are discussed and shown to have acceptable complexity in practice. It is expected that these strategies could also be applied in our scenario.

4. Dynamic Extractor

Using the MPEG-21 DIA descriptions, the ADTE determines the optimal amount of discarded bitrate $\{p_i\}$. This information is then pushed to the extractor to carry out the adaptation at NAL unit level. The procedure inside our extractor is essentially based on the method of [4].

As shown in Figure 6, the dynamic extractor consists of two key sub-modules, the discarder and the layer dependency modifier (LDM). The discarder sequentially discards MGS NAL units in spatial layer i according to the decided value p_i . However, for lower spatial layers, the NAL units containing motion vector data will never be discarded, so that inter-layer motion prediction is not

broken. Further, the headers of retained NAL units may be modified by the LDM to make the adapted bitstream syntactically compliant. Detailed explanation of the modification procedure can be found in [4]. Further, it is shown in [4] that the processing time of the extractor is negligible. It should be noted that any adaptation option that drops one or more NAL unit in a lower spatial layer would need this procedure.

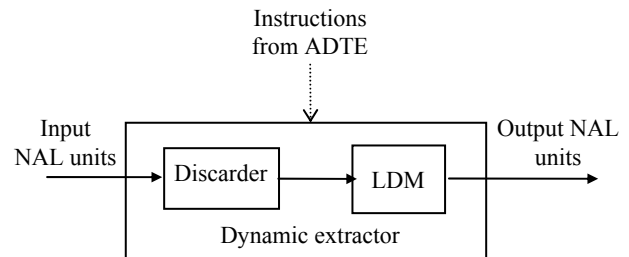


Fig. 6. Sub-modules of the dynamic extractor

In some MPEG-21 related work, XML-based bitstream syntax description (BSD) is used to carry out the actual adaptation of SVC bitstreams. However, the reported processing time is still rather high due to the complexity in XML processing [19]. As SVC standard already provides a good system interface (i.e. high-level syntax) [11], one more wrapper (here BSD) may be unnecessary for simple adaptation operations of SVC.

III. EXPERIMENTS

This section provides experiment results that show the effectiveness of the proposed approach. In our system, the encoder, extractor, and decoder are based on SVC reference software JSVM9.11. Each bitstream is encoded with 2 spatial layers, namely QCIF and CIF, both having frame rate of 30fps. The GOP size is 16 and I-picture period is 32 to enable random access to the content. Base quantization parameters of the bottom and top spatial layers are 38 and 42 respectively. Each spatial layer is enhanced by 2 MGS stacks, and each stack is composed of 7 MGS layers. The MGSVector of each MGS stack is [0, 1, 1, 2, 3, 4, 5]. That means the first MGS layer of an MGS stack contains no residual data. By this way, most MGS NAL units (except those containing motion information) of a lower spatial layer could be discarded. Quality metric used in the experiments is average PSNR. Suppose that two users will consume an adapted bitstream as in the above scenario. The AdaptationQoS modules in Figure 4 and Figure 5 are the R-D data of the Football test sequence, with discarding step of 100Kbps.

Figure 7 compares the quality of three adaptation

options bottom-max ($w_1=0, w_2=1$), top-max ($w_1=1, w_2=0$), and a trade-off option with $w_1=0.33$ and $w_2=0.67$. The quality curves of top layer (Figure 7a) and the bottom layer (Figure 7b) are shown with respect to the total bitrate of the adapted bitstream. So the curves of the top layer are the actual R-D curves (but it is not the case with the bottom layer). Also, as the gain of the top-max adaptation compared to bottom-max is significant chiefly at low bitrates [4], the results here are shown in low bitrate range for better visibility. It can be seen that top-max and bottom-max options are the two extreme cases of quality optimization; one spatial layer is maximized but the other layer is degraded. Meanwhile, the trade-off curve provides a balance between the two extremes. For example, at the total bitrate of 1030kbps the quality in the top layer provided by trade-off option is about 0.54dB lower than that of the top-max option; however, in the bottom layer, the quality of trade-off option is 1.77dB higher than that of the top-max option.

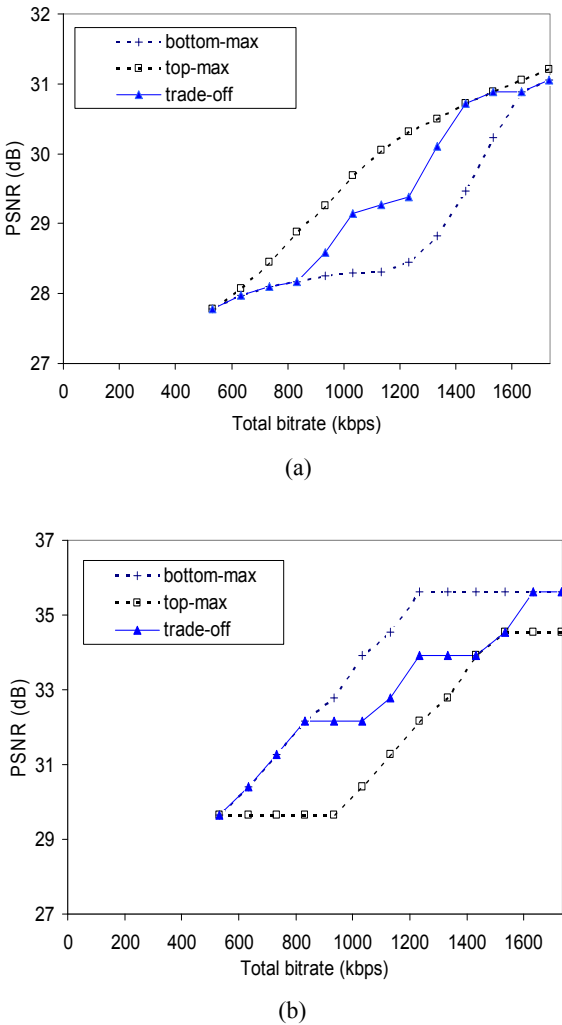


Fig. 7. Comparison of different adaptation options for Football sequence. The trade-off curve has $w_1=0.33, w_2=0.67$. (a) Quality of the top layer. (b) Quality of the bottom layer.

Figure 8 shows the trade-off option with $w_1=0.2, w_2=0.8$. Compared to previous values, these weight values emphasize somewhat the top layer. As a result, in Fig. 8 the quality curves of the trade-off option move closer to the top-max curves. So, the trade-off between the two extreme cases can be flexibly controlled by adjusting the weight values.

In Figure 9, similar results are shown for the News sequence. Two trade-off options, one with ($w_1=0.5, w_2=0.5$) and the other with ($w_1=0.6, w_2=0.4$), are shown in the same graph. With this bitstream, the option with ($w_1=0.5, w_2=0.5$) gives a good balance between the top-max and bottom-max option.

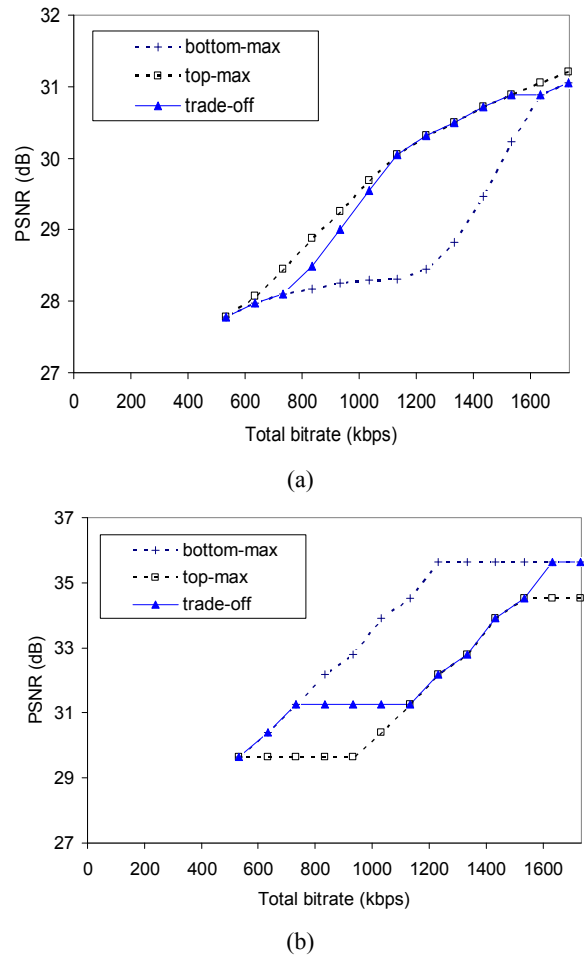


Fig. 8. Comparison of different adaptation options for Football sequence. The trade-off curve has $w_1=0.2, w_2=0.8$. (a) Quality of the top layer. (b) Quality of the bottom layer.

It can be seen that the gap between top-max and bottom-max can be quite large, thus the range of tradeoff between layers is rather wide. With the above selected trade-off options, the improvement compared to an extreme option is up to several dBs. Of course, in our approach 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. The above results also show that good values of weights depend on the bitstream itself. Moreover, it can be expected that the weights will also depend on the quality metrics used. In practice, some weights may be computed in advance and stored in the same way as AdaptationQoS description.

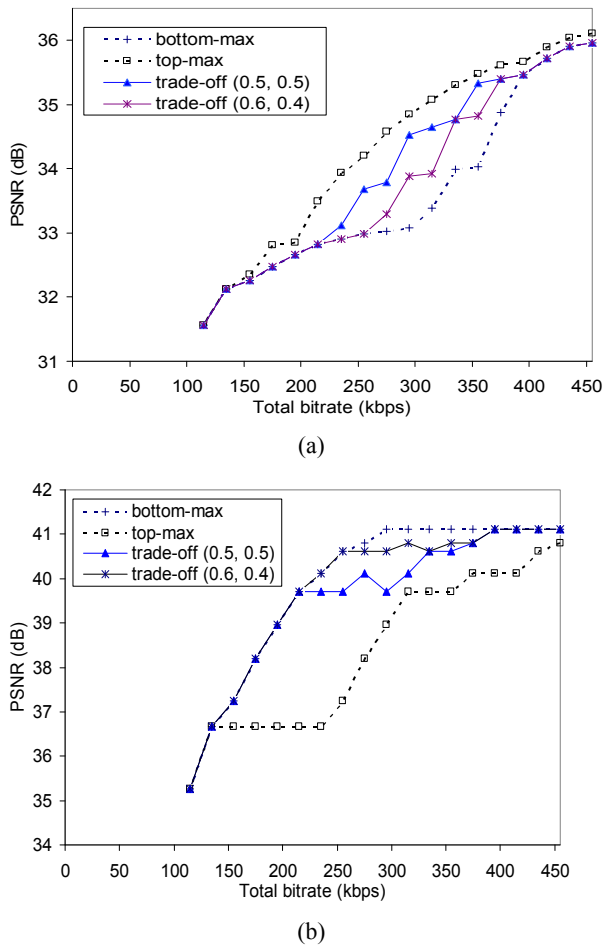


Fig. 9. Comparison of different adaptation options for News sequence. (a) Quality of the top layer. (b) Quality of the bottom layer.

As MGS is packet-based scalability, there could be a doubt that the actual adapted bitrate may widely deviate from the requested bitrate. However, it turns out that the actual adapted bitrate matches well the requested bitrate. For example, in Figure 10, the nodes on each curve represent a series of bitrates which are created by repeatedly discarding 50Kbps in the top spatial layer. Here, each curve corresponds to an amount of discarded (or truncated) bitrate in the bottom layer (denoted as R0). We

can see that the actual adapted bitrates well reflect the (requested) amounts of discarded data. This can be explained that though discarding a packet (i.e. a NAL unit) is not as "fine-grained" as in FGS, the discarded data size of the packet is actually spread over some time interval thus adapted bitrate variations are rather small. So, with an MGS-based bitstream, it is possible to get nearly any adapted bitrate within the supported bitrate range.

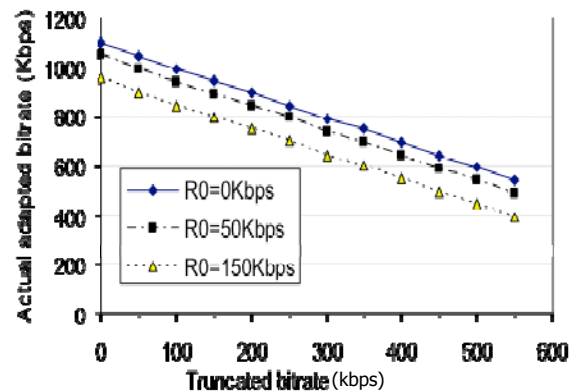


Fig. 10. Actual adapted bitrate w.r.t discarded bitrate.

As mentioned above, the processing time of both ADTE and extractor is negligible. 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 adapted accordingly. That means the whole adaptation process is seamless to the users.

V. CONCLUSIONS

In this paper, we have studied the adaptation of SVC bitstream for multiple users. Our approach was shown to be standard-compliant through the use of the MPEG-21 multimedia framework and a conforming SVC bitstream at the output. Especially, with a controllable tradeoff between spatial layers, our approach could harmonize the qualities of different user groups in an optimal manner. The experiment results showed that our system was effective to support flexible and interoperable adaptation in heterogeneous environments. Our future work will focus on adaptation behavior of SVC video which is guided by perceptual quality metrics.

REFERENCES

- [1] C. Singhal, S. De, R. Trestian, and G. M. Muntean, "Joint optimization of user-experience and energy-efficiency in wireless multimedia broadcast," *IEEE*

- Trans. Mobile Computing*, vol. 13, no. 7, pp. 1522–1535, October 2014.
- [2] D. S. Kim, S. Y. Lee, S. S. Lee, J. H. Ahn, and B. H. Cho, "Implementation of a H.264/AVC SVC decoder with multi-symbol prediction CAVLC for advanced T-DMB receiver", *IEEE Trans. Consumer Electron.*, vol. 56, no. 4, pp. 2819 - 2825, Nov. 2010.
- [3] H. Schwarz, D. Marpe, and 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.
- [4] T. C. Thang, J. W. Kang, J. J. Yoo, and J. G. Kim, "Multilayer Adaptation for MGS-based SVC Bitstream", *Proc. ACM Multimedia 2008*, pp. 689-692, Oct. 2008.
- [5] T. C. Thang, Y. S. Kim, Y.M. Ro, J. W. Kang, J.-G. Kim, "SVC Bitstream Adaptation in MPEG-21 Multimedia Framework", *Journal of Zhejiang University SCIENCE A*, vol. 7, no. 5, pp. 764-772, May 2006.
- [6] E.-D. Jang, J.-G. Kim, T. C. Thang, and J. W. Kang, "Adaptation of Scalable Video Coding to Packet Loss and its Performance Analysis", *Proc. ICACT2010, Seoul*, Feb. 2010.
- [7] H. Lee, Y. Lee, J. Lee, and D. Lee, Heonshik Shin, "Design of a Mobile Video Streaming System Using Adaptive Spatial Resolution Control", *IEEE Trans. Consumer Electron.*, vol. 55, no. 3, pp. 1682 – 1689, Aug. 2009.
- [8] H. Zhang, H. Nguyen, E. M. Garcia, P. A. T. Solano, D. Zhang, N. Crespi and B. Guo, "Scalable multimedia delivery with QoS management in pervasive computing environment", *Journal of Supercomputing*, vol. 65, pp. 317-335, Jul. 2013.
- [9] T. C. Thang, J. W. Kang, N. Pham-Ngoc, and A. T. Pham, "Standard-Compliant Content Adaptation in IPTV Systems" (Invited paper), *REV Journal on Electronics and Comm.*, vol. 2, no. 1-2, pp. 26-32, June, 2012.
- [10] D. De Schrijver, W. De Neve, K. De Wolf, R. De Sutter, and R. Van de Walle, "An optimized MPEG-21 BSDL framework for the adaptation of scalable bitstreams," *Journal of Visual Comm. and Image Representation*, vol. 18, no. 3, pp. 217-239, June 2007.
- [11] T. C. Thang, J. G. Kim, J. W. Kang, and J. J. Yoo, "SVC Adaptation: Standard Tools and Supporting Methods", *Signal Processing: Image Communication*, vol. 24, no. 3, pp.214-228, Mar. 2009.
- [12] T. C. Thang, J. W. Kang, J. J. Yoo, and Y. M. Ro, "Optimal multi-layer adaptation of SVC video over heterogeneous environments", *Journal of Advances in Multimedia*, vol. 2008, Article ID 739192, 8 pages, 2008.
- [13] 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, June 2005.
- [14] I. Burnett, F. Pereira, R. Van de Walle, and R. Koenen, *The MPEG-21 Book*, John Wiley & Sons, 2006.
- [15] L. Lin, A. Ortega, "Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 4, pp. 446-459, Aug. 1998.
- [16] G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268-278, March 1973.
- [17] T. C. Thang, Y. J. Jung, and Y. M. Ro, "Effective Adaptation of Multimedia Documents with Modality Conversion," *Signal Processing: Image Communication*, vol. 20, no.5, pp. 413-434, May 2005.
- [18] I. Kofler, C. Timmerer, H. Hellwagner, A. Hutter, and F. Sanahuja, "Efficient MPEG-21-based Adaptation Decision-Taking for Scalable Multimedia Content", *Proc. MMCN'07*, San Jose, Jan. 2007.
- [19] D. Van Deursen, W. De Neve, D. De Schrijver, and R. Van de Walle, "BFlavor: An Optimized XML-based Framework For Multimedia Content Customization," *Proc. 25th Picture Coding Symposium*, Beijing, 2006.

AUTHORS



Truong Cong Thang received the B.E. degree from Hanoi University of Technology, Vietnam, in 1997 and Ph.D. degree from KAIST, Korea, in 2006. From 1997 to 2000, he worked as an engineer in Vietnam Post & Telecom (VNPT). From 2007 to 2011, he was a Member of Research Staff at Electronics and Telecommunications Research Institute (ETRI), Korea. He has been an active member of Korean and Japanese delegations to standard meetings of ISO/IEC and ITU-T from 2002. Since 2011, he has been an Associate Professor of the University of Aizu, Japan. His research interests include multimedia networking, image/video processing, content adaptation, IPTV, and MPEG/ITU standards.



Hung T. Le received the B.E. degree from Hanoi University of Technology, Vietnam, in 2012 and M.E. degree from University of Aizu, Japan, in 2014. Since 2012, he has been a research assistant at

the Computer Communications Lab., the University of Aizu. His major research interests include multimedia networking, adaptive video streaming, QoE support. He is a student member of IEEE.



Duc V. Nguyen received the B.E. degree from the University of Aizu, Japan, in Mar. 2014. He is currently a research assistant in the Computer Comm. Lab., the University of Aizu. His research interests include multimedia networking, content adaptation, and IPTV. He received the 3rd place of the best student paper award in the 2015 IEEE Conf. on Consumer Electronics (ICCE2015).



Anh T. Pham received the B.E. and M.E. degrees, both in Electronics Engineering from the Hanoi University of Technology, Vietnam in 1997 and 2000, respectively, and the Ph.D. degree in Information and Mathematical Sciences from Saitama University, Japan in 2005. From 1998 to 2002, he was with the NTT Corp. in Vietnam. Since April 2005, he has been on the faculty at the University of Aizu, where he is currently a Professor of Computer Communications with the Division of Computer Engineering. He also holds an adjunct professor position at Vietnam National University/University of Engineering and Technology. Prof. Pham's research interests are in the broad areas of communication theory and networking with a particular emphasis on modeling, design and performance evaluation of wired/wireless communication systems and networks. He received Japanese government scholarship (MonbuKagaku-sho) for PhD study. He also received Vietnamese government scholarship for undergraduate study. Prof. Pham is senior member of IEEE. He is also member of IEICE and OSA.

This is blank Page