Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

SVC bitstream adaptation in MPEG-21 multimedia framework

THANG Truong Cong[1], KIM Young Suk[1], RO Yong Man[‡1], KANG Jungwon[2], KIM Jae-Gon[2]

([1]*Department of Engineering, Information and Communications University, Daejeon 305-732, Korea*)

([2]*Electronics and Telecommunications Research Institute, Daejeon 305-732, Korea*)

E-mail: {tcthang; coin0104; yro}@icu.ac.kr; {jungwon; jgkim}@etri.re.kr

Received Dec. 15, 2005; revision accepted Feb. 20, 2006

Abstract: Scalable video coding (SVC) is the most promising video format for applications of collaborative communication. MPEG-21 standard has newly emerged to enable the interoperability of multimedia delivery in heterogeneous environments. In this paper we study adaptation of SVC bitstream in the context of MPEG-21 multimedia framework. For interfacing SVC bitstream with MPEG-21 based adaptation system, we propose three SVC specific adaptation operators. Based on our previous work with multidimensional video adaptation, we present an effective approach, using MPEG-21 DIA AdaptationQoS description tool, to model QoS control for SVC adaptation. Then we show how the operator values could be computed from that representation. For the actual adaptation at bitstream level, we propose a procedure to remove the unnecessary NAL units from an SVC bitstream. The result of this study enables QoS management of SVC streaming in an efficient and standardized manner.

Key words: Multidimensional video adaptation, Scalable video coding (SVC), MPEG-21 DIA (Digital Item Adaptation), QoS
doi:10.1631/jzus.2006.A0764 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

Scalable video coding (SVC) (Schwarz *et al*., 2004) is the most promising video format for applications of collaborative communication. First, SVC format, which is extended from the latest advanced video coding (AVC), is appropriate for creating a wide variety of bitrates with high compression efficiency. Second, an original SVC bitstream can be easily truncated in different ways to meet various characteristics and variations of devices and connections. The scalability is possible in 3 dimensions: spatial, temporal, and SNR. The MPEG-21 standard has newly emerged to enable the interoperability of multimedia delivery in heterogeneous environments. Especially, part 7 of MPEG-21 standard (DIA—Digital Item Adaptation) includes the description tools for adapting multimedia contents (Vetro, 2004).

In this work, we study the use of MPEG-21 DIA AdaptationQoS description tool to adapt SVC bitstream. AdaptationQoS describes relations between resource requirements (e.g. bitrate), adaptation operations, and the corresponding qualities. Adaptation operators for various content types/formats are provided in the classification schemes (CS) of MPEG-21 DIA (ISO/IEC IS 15938-5:2001, 2003). As SVC is a new video coding format, we propose three adaptation operators specific to the structure of SVC bitstream. The first operator is SpatialLayers, which indicates the number of spatial enhancement layers to be discarded from the input SVC bitstream. The second operator is TemporalLevels, which indicates the number of temporal enhancement layers to be discarded. And the third is QualityReduction, which indicates the fraction of SNR enhancement layers to be truncated from the input bitstream.

Based on our previous work with multidimensional video adaptation (Jung *et al*., 2005a), we present an effective approach, using AdaptationQoS description tool, to model QoS control for SVC adaptation. Specifically an adaptation trajectory can be

‡ Corresponding author

represented by AdaptationQoS's utility function having a finite number of discrete points. Then we show how the values of adaptation operators, at any available connection bitrate, can be computed from the points of the utility function. We also propose an efficient procedure for adapting the SVC bitstream given the three operator values. For experiment, we develop an adaptation system that includes a decision engine (which employs DIA description tools) and a scaling engine (which carries out the adaptation procedure). The usefulness of our adaptation system is shown through some adaptation examples.

This paper is organized as follows. In Section 2, we first describe the general concept of multidimensional video adaptation based on user preference, then we briefly review the basics of SVC. In Section 3, we propose three new SVC-specific operators together with an adaptation procedure in the context of MPEG-21 DIA framework. The experiment is presented in Section 4 and finally conclusions and future work are provided in Section 5.

MULTIDIMENTIONAL VIDEO ADAPTATION

User preference based adaptation

Video adaptation is an important technique to provide video QoS management in heterogeneous networks (Chang and Vetro, 2005). Basically, video can be adapted in three dimensions as follows:

(1) SNR/quality dimension: the visual clarity of each picture is reduced (scaled).

(2) Temporal dimension: video is scaled by reducing frame rate (temporal resolution).

(3) Spatial dimension: video is scaled by reducing spatial size (spatial resolution).

To provide the best quality to user, the tradeoff between these dimensions should be considered. For example, at a given bitrate constraint, the provider should know whether an adapted video with low spatial resolution and high frame rate is better than another one having the same bitrate but low frame rate and high spatial resolution.

In video adaptation, quality measure/utility is important to guide the adaptation process, more specifically to find the optimal tradeoff between the adaptation dimensions. However, currently quality measures for multi-dimensional (spatial, temporal, SNR) video adaptation are not mature yet. Most current researches devise some rather ad-hoc quality measure/criteria to guide the adaptation process (Jung and Ro, 2004; Jung *et al*., 2005b).

In practice, video streaming service over mobile networks often provides short video clips to users in different genres. It is noted that video clips of the same genre (or sub-genre) have similar adaptation behavior or adaptation trajectory/path. We carried out some extensive subjective tests to identify the adaptation characteristics of common genres in mobile video service (Jung *et al*., 2005a). At a given bitrate, each original video is adapted into a number of bitrate-equivalent versions; each version has different temporal/spatial/SNR characteristics. The users will select the best version for a bitrate (forced choice method). The set of best versions in a wide bitrate range constitutes an adaptation trajectory for the original video.

Fig.1 shows two adaptation trajectories, in 3D space, for two genres (or concepts): crowd and text & graphics (T & G). Here, video may have two spatial size/resolutions (CIF and QCIF) and three frame rates (30, 15, 7.5). In Fig.1, each adaptation trajectory is composed of a set of segments $\{O_iP_i\}$. Note that bitrate of P_i is equal to bitrate of O_{i+1}. With a new O_iP_i segment, there is a change in frame rate (temporal resolution) and/or frame size (spatial resolution) compared to the previous segment. And on a segment, the bitrate is reduced only by degrading SNR quality. We call $\{P_i\}$ the set of "switching points".

With crowd genre, we see that when the bitrate is reduced, the user first wants to keep the original spatial and temporal resolutions, i.e. CIF@30fps, and reduce just the SNR quality (segment O_1P_1). As bitrate is further reduced, the user wants to keep original spatial resolution but reduce temporal resolution, i.e. CIF@15fps in segment O_2P_2. The process continues similarly in other segments. With the genre text & graphics (T & G), the user just wants to change the temporal resolution and SNR quality, but not the spatial resolution.

The adaptation trajectories of different contents in the same genre are rather consistent. This can be seen in Fig.2, where we show the relative bitrates of switching points (w.r.t. corresponding original bitrates) for various video clips in the genre crowd. More details of adaptation trajectory can be found in (Jung *et al*., 2005a).

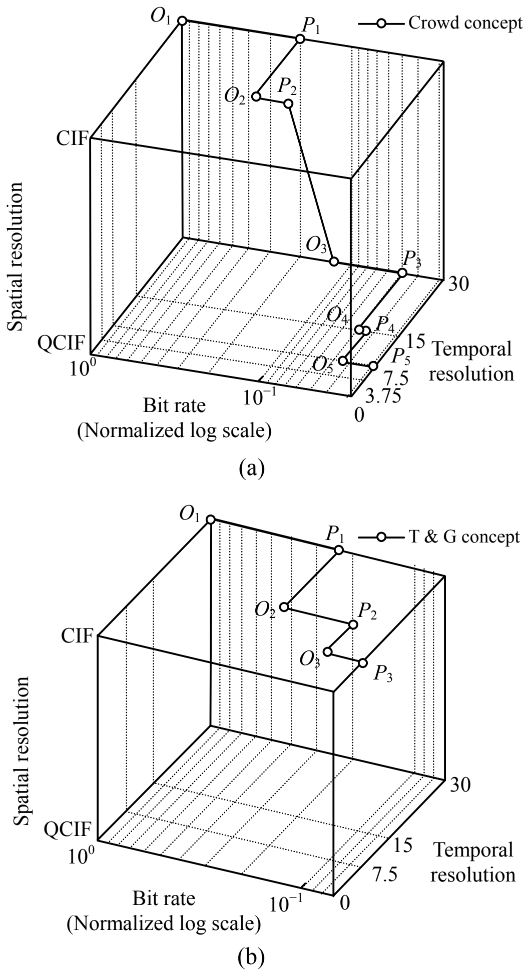

Fig.1 Adaptation trajectory of (a) crowd genre and (b) text & graphics genre

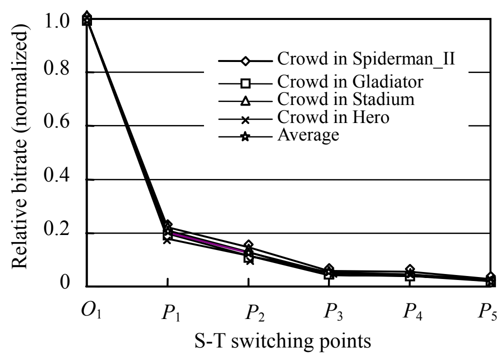

Fig.2 Relative bitrates at switching points for different video clips in crowd genre

The above concept of adaptation trajectory can be used to build practical AdaptationQoS descriptions for video adaptation as shown in Section 3. In the following, we present a brief introduction to SVC.

Scalable video coding

Video adaptation based on transcoding is notoriously complex and time-consuming. Scalable Video Coding (SVC), which is being jointly developed by MPEG and VCEG, aims at the 3D scalability by simple truncation of appropriate parts in bitstreams (JSVM 3.0, 2005). Fundamentally, there are two different ways to support scalability, either by using a technique that is intrinsically scalable (e.g. bitplane arithmetic coding) or by using a layered approach. In SVC, a combination of the two approaches to enable a full spatio-temporal and quality scalability is used.

Temporal scalability is enabled by hierarchical prediction (hierarchical B-frames or MCTF), whereas spatial scalability is provided using a layered approach. For quality (SNR) scalability, two different possibilities are provided; an embedded quantization approach for coarse grain scalability (CGS) and a fine grain scalability (FGS) approach based on the principle of sub-bitplane arithmetic coding. In SVC, a CGS layer is coded in the same way as a spatial enhancement layer. To obtain high coding efficiency, the texture, residual, and motion vectors of high spatial layer is predicted from low spatial layer. To support backward compatibility, the base quality layer (i.e. without SNR enhancement) of the lowest spatial layer is coded by AVC format.

The illustration of an example SVC bitstream is shown in Fig.3. This bitstream contains two spatial layers: QCIF encoded at 15 fps and CIF encoded at 30 fps. Each spatial layer is composed of one base quality layer and one SNR enhancement layer. The pictures at the middle row are temporary prediction

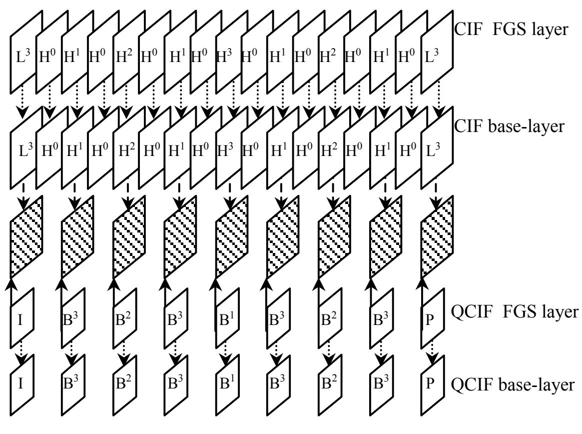

Fig.3 Illustration of an SVC bitstream

of CIF layer. In temporal dimension, each picture belongs to one temporal layer indicated by its subscript number.

Similar to AVC, an SVC bitstream is divided into NAL (network abstraction layer) units. Except NAL units of AVC base layer, all other NAL units have three key parameters for decodability dependency information in their header, namely dependency_id, temporal_level, and quality_level.

Essentially, dependency_id indicates the identifier of a spatial layer or a CGS layer; temporal_level indicates the level of a temporal layer; and quality_level indicates the identifier of an FGS layer. quality_level equal to 0 represents a spatio-temporal base layer (with a certain value of dependency_id and temporal_level). The NAL unit header plays a key role in extracting/adapting SVC bitstream. Any layers in SVC bitstream can be removed simply by discarding the NAL units having appropriate headers. Further, the NAL units of FGS layer can be truncated arbitrarily to refine the bitrate of adapted bitstream.

SVC BITSTREAM ADAPTATION WITH MPEG-21 DIA

Overview

A typical architecture of an adaptation system is depicted in Fig.4. An adaptation system basically contains a decision engine and a scaling engine (Thang *et al.*, 2005b). The decision engine employs AdaptationQoS description and other metadata like connection bitrate, bitstream information, and then decides adaptation instructions (specifically, values of operators in our system). The scaling engine takes the instructions and adapt the input bitstream accordingly.

Note that the description tools to support the

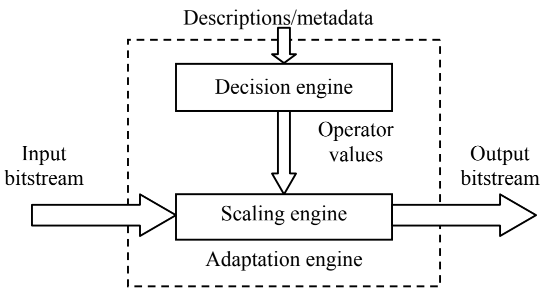

Fig.4 Adaptation engine architecture

processing of adaptation engine is scope of MPEG-21 DIA and the decoder is scope of SVC.

In the following, we first propose SVC adaptation operators in MPEG-21 DIA, then show how the values of operators can be decided using an AdaptationQoS description, and finally present a procedure to adapt the SVC bitstream given the operator values.

Proposed operators for MPEG-21 DIA AdaptationQoS

The MPEG-21 standard has newly emerged to enable the interoperability of multimedia delivery in heterogeneous environments. Part 7 of MPEG-21 standard (DIA—Digital Item Adaptation) includes the description tools to facilitate adaptation of multimedia contents. Among these tools, AdaptationQoS is one important tool to manage QoS at application level. This tool describes relations between resource requirements (e.g. bitrate), adaptation operations, and the corresponding qualities. Adaptation operators for various content types/formats are provided in the classification schemes (CS) of MPEG-21 DIA.

Though AdaptationQoS tool provides the needed information or parameters for many different kinds of media resources, SVC bitstream is not considered yet. With current adaptation operators in DIA such as P-frame dropping, coefficient dropping, requantization, etc., it is not proper for supporting adaptation for SVC bitstream.

To support the adaptation of SVC bitstream, we propose the SVC adaptation operators in MPEG-21 DIA as described below (Thang *et al.*, 2005c).

(1) SpatialLayers: indicates the number of enhancement layers for spatial resolution to be truncated from the input bitstream. It is assumed that the highest enhancement layer is truncated first.

(2) TemporalLevels: indicates the number of enhancement layers for temporal resolution to be truncated from the input bitstream. It is assumed that the highest enhancement layer is truncated first.

(3) QualityReduction: indicates the SNR enhancement fraction that should be truncated from the input bitstream. It is assumed that the Quality Reduction for no truncation is “0” and for full truncation is “1”.

Decision on operator values

First, based on the concept of adaptation trajectory and the above adaptation operators, we present a simple but efficient method to build AdaptationQoS description for video contents. A description will contain point O_1 and all points P_i's of an adaptation trajectory. An example of AdaptationQoS description for a content of text & graphics genre is shown in Fig.5. This description contains four points, each point is associated with one bitrate (constraint) value, three operator values, and one quality (utility) value. The adaptation trajectories above are obtained from forced choice methodology, so the quality value of each adaptation point along a trajectory is just a relative value. In this case, the utility rank of AdaptationQoS is used to describe the quality ordering of adaptation points. Note that the AdaptationQoS description in this example is in the form of a utility function, although the adaptation trajectory can be similarly represented in the form of lookup table. More information about using AdaptationQoS can be found in (ISO/IEC IS 15938-5:2001, 2003; Mukherjee *et al*., 2005).


```
<Description xsi:type="AdaptationQoSType">
  <Module xsi:type="UtilityFunctionType">
    <Constraint iOPinRef="BANDWIDTH">
        <Values xsi:type="FloatVectorType">
          <Vector>75 355 789 1546</Vector>
        </Values>
    </Constraint>
    <AdaptationOperator  iOPinRef="SPATIALLAYERS">
        <Values xsi:type="IntegerVectorType">
          <Vector>0 0 0 0</Vector>
        </Values>
    </AdaptationOperator>
    <AdaptationOperator iOPinRef="TEMPORALLEVELS">
        <Values xsi:type="IntegerVectorType">
          <Vector>2 1 0 0</Vector>
        </Values>
    </AdaptationOperator>
    <AdaptationOperator iOPinRef="QUALITYREDUCTION">
        <Values xsi:type="FloatVectorType">
          <Vector>0.99 0.71 0.28 0</Vector>
        </Values>
    </AdaptationOperator>
    <UtilityRank>1 2 3 4</UtilityRank>
  </Module>
  <IOPin id="BANDWIDTH"/>
  <IOPin id="SPATIALLAYERS"/>
  <IOPin id="TEMPORALLEVELS"/>
  <IOPin id="QUALITYREDUCTION"/>
</Description>
```


Fig.5 Example of AdaptationQoS description

AdaptationQoS description as built above consists of the key points only (O_1 and switching points). So the number of points used in the description is very small and description can be automatically created based on adaptation trajectory.

Based on AdaptationQoS description, we need to compute the values of operators for a given connection bitrate. Obviously when the connection bitrate is equal to the bitrate (constraint) of a key point of utility function, the decision engine can immediately output the values of three operators associated with that point.

For some operators in MPEG-21 DIA, when the connection bitrate is arbitrary, the operator values may be computed by interpolating the neighboring key points. However, the SVC adaptation operators are different and can be computed as follows.

Denote $B_{\mathrm{o}}^{\mathrm{SNR}}$ the original bitrate of all SNR enhancement layers, B_{c} the connection bitrate, and P_{a} the key point of utility function whose bitrate $B_{P_{\mathrm{a}}}$ is smaller than and nearest to B_{c}. We see that the adapted content will correspond to some position on the segment $Q_{\mathrm{a}}P_{\mathrm{a}}$. Also denote $SpatialLayers_{P_{\mathrm{a}}}$, $TemporalLevels_{P_{\mathrm{a}}}$, $QualityReduction_{P_{\mathrm{a}}}$ as the operator values at point P_{a}. We have the output values of the first two operators given by:

$$SpatialLayers = SpatialLayers_{P_{\mathrm{a}}}, \tag{1}$$

$$TemporalLevels = TemporalLevels_{P_{\mathrm{a}}}. \tag{2}$$

As for *QualityReduction*, first we see that bitrate $B_{P_{\mathrm{a}}}^{\mathrm{SNR}}$ of SNR enhancement layers at point P_{a} is:

$$B_{P_{\mathrm{a}}}^{\mathrm{SNR}} = (1 - QualityReduction_{P_{\mathrm{a}}}) \cdot B_{\mathrm{o}}^{\mathrm{SNR}}. \tag{3}$$

Then the bitrate $B_{\mathrm{c}}^{\mathrm{SNR}}$ of SNR enhancement layers at the position corresponding to B_{c} is

$$B_{\mathrm{c}}^{\mathrm{SNR}} = B_{P_{\mathrm{a}}}^{\mathrm{SNR}} + B_{\mathrm{c}} - B_{P_{\mathrm{a}}}. \tag{4}$$

So the output *QualityReduction* is computed by

$$\begin{aligned} QualityReduction &= \frac{B_{\mathrm{o}}^{\mathrm{SNR}} - B_{\mathrm{c}}^{\mathrm{SNR}}}{B_{\mathrm{o}}^{\mathrm{SNR}}} \\ &= \frac{(B_{\mathrm{o}}^{\mathrm{SNR}} \cdot QualityReduction_{P_{\mathrm{a}}} - B_{\mathrm{c}} + B_{P_{\mathrm{a}}})}{B_{\mathrm{o}}^{\mathrm{SNR}}}. \end{aligned} \tag{5}$$

It should be noted that although adaptation operators provided by MPEG-21 DIA CSs are normative, applications can use proprietary or third party CSs. That means the operators themselves are also topics of research; new operators can be devised to improve the adaptation effectiveness (e.g., for some specific contexts).

Adaptation procedure

Given the three values of operators, the scaling engine needs to accordingly adapt/scale the input bitstream. First, we have the following definitions:

(1) Effective spatial layer SL_e: is the index of a spatial layer of SVC bitstream. The lowest spatial (base) layer has SL_e equal to 0, and for each higher spatial enhancement layer, its value of SL_e is increased by one.

(2) Effective temporal level TL_e: is the index of a temporal layer. The lowest temporal layer has TL_e equal to 0, and TL_e of each higher temporal enhancement layer is increased by one.

Given an SVC bitstream, the bitstream analysis process will provide the mapping SL_e (dependency_id) which shows the corresponding value of SL_e of a dependency_id. AVC NAL units do not have parameter dependency_id, however, there is only one base AVC layer, so its SL_e can be easily set to 0. Meanwhile, each CGS layer also has a unique dependency_id, but CGS layer is for SNR enhancement, not spatial (resolution) enhancement. So, SL_e of a CGS layer's dependency_id will be equal to SL_e of the spatial base layer enhanced by that CGS layer.

Essentially, each effective temporal level corresponds to one value of frame rate supported by the SVC bitstream. The analysis process will provide the mapping TL_e (temporal_level) which shows the corresponding value of TL_e of a temporal_level. As NAL unit headers of AVC layer do not have temporal_level parameter, the value of TL_e is derived from the parameter nal_ref_idc of NAL unit header. For example, suppose we have an SVC bitstream with two spatial layers QCIF@15fps and CIF@30fps and the GOP size is 16 frames. The CIF layer will have five (effective) temporal levels (0~4), and the QCIF layer (i.e. AVC) has two effective temporal levels which are 2 (when *nal_ref_idc*≠3) and 3 (when *nal_ref_idc*=3).

Denote SL_e^{max} and TL_e^{max} as the maximum values of SL_e and TL_e, again B_o^{SNR} as the bitrate of SNR enhancement layers in the original SVC bitstream, and B_b^{SNR} as the bitrate budget for SNR enhancement layers in adapted SVC bitstream.

Then given three values of *SpatialLayers*, *TemporalLevels*, and *QualityReduction*, the procedure to adapt SVC bitstream is as follows:

Step 1: Removal of spatial enhancement layers

In this step, all NAL units having $SL_e > (SL_e^{max} - SpatialLayers)$ are removed (i.e., quality base layers as well as associated FGS/CGS layers).

Step 2: Removal of temporal enhancement layers

In this step, all NAL units having $TL_e > (TL_e^{max} - TemporalLevels)$ are removed (i.e., quality base layers as well as associated FGS/CGS layers).

Step3: Truncation of SNR enhancement layers

In this step, the SNR enhancement layers of the remaining spatial and temporal layers (both base and enhancement) are truncated.

Denote B_r as the amount of FGS/CGS bits removed in the above two steps. We have the bit budget for SNR enhancement layers as follows:

$$B_b^{SNR} = (1 - QualityReduction) B_o^{SNR}. \quad (6)$$

Also we denote:

$NUM_{SNR}[i]$: the number of SNR enhancement layers (both FGS and CGS) of effective spatial layer i;

$SNRen\_lay[i, j]$: SNR enhancement layer j at effective spatial layer i;

B_{acu}: accumulative bitrate of SNR enhancement layers.

$Bit(SNRen\_lay[i, j])$: bitrate of $SNRen\_lay[i, j]$.

The truncation of SNR enhancement layers is carried out by the following sub-procedure:

```
Bacu=0;
  for (i=0; i≤( SLe^max −SpatialLayers); i++) {
    for (j=0; j<NUMSNR[i]; j++) {
      if (Bit(SNRen_lay[i, j])+Bacu< Bb^SNR ) {
        SNRen_layer[i, j] is included;
        Bacu+=Bit(SNRen_lay[i, j])
      }
      else {
        if ((SNRen_layer[i, j] is a CGS layer) and
              (i==( SLe^max −SpatialLayers)))
          SNRen_layer[i, j] is removed;
        else {  // it is an FGS layer
          if (Bacu< Bb^SNR )
```

SNRen_layer[*i*, *j*] is truncated to ($B_b^{SNR} - B_{acu}$);
else
SNRen_layer[*i*, *j*] is removed;
}
$B_{acu} = B_b^{SNR}$;
}
}
}

Here, when an SNR enhancement layer is removed (or included), all NAL units of that layer are removed (or included). And when an SNR enhancement layer is truncated, all NAL units of that layer are truncated with the same ratio.

In the above procedure, although FGS and CGS layers have different coding modes and different *dependency_id*'s, we can still treat them in the same manner by using the "effective spatial layer". The difference is that CGS layer should be wholly included or wholly removed while FGS can be partially truncated. Moreover, a CGS layer can be removed only if it is for enhancement of the highest spatial layer. The signaling of CGS/FGS layer can be derived from some metadata describing SVC bitstream, e.g., scalability information SEI message (Wang and Thang, 2005).

EXPERIMENT

Adaptation experiment

In this part, we show an experiment on SVC bitstream adaptation. We extend the adaptation system developed in (Jung *et al.*, 2003) to support SVC adaptation. The SVC scaling engine is built based on the extractor of JSVM2.0.

We employ the Football video sequence in this experiment. The original video is encoded with two spatial layers: QCIF@30fps and CIF@30fps. Each spatial layer is enhanced by 3 FGS layers. GOP size is 16, and video length is 5 s. Total bitrate of original content is 2400 kbps, and the bitrate of all FGS layers is 2179 kbps.

The adaptation trajectory of the Football video is shown in Fig.6. We can see that when the bitrate is reduced, the adaptation will follow segment O_1P_1 (CIF@30fps), then O_2P_2 (CIF@15fps), O_3P_3 (QCIF @30fps), and finally O_4P_4 (QCIF@15fps).

Based on the adaptation trajectory, the AdaptationQoS description is created in the form of a utility function consisting of five key points O_1, P_1, P_2, P_3, and P_4 as shown in Fig.7.

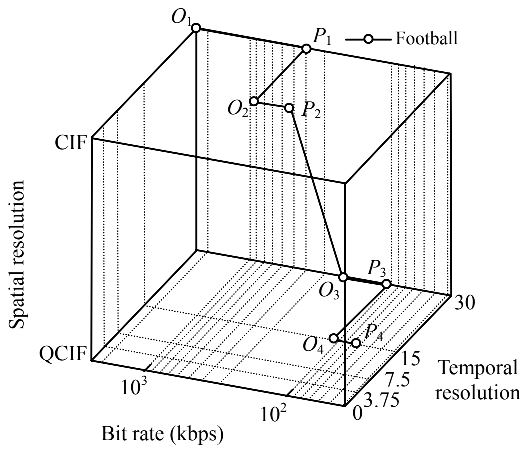

Fig.6 Adaptation trajectory for Football video

```
<Description xsi:type="AdaptationQoSType">
  <Module xsi:type="UtilityFunctionType">
    <Constraint iOPinRef="BANDWIDTH">
      <Values xsi:type="FloatVectorType">
        <Vector>77 110 220 400 2400</Vector>
      </Values>
    </Constraint>
    <AdaptationOperator iOPinRef="SPATIALLAYERS">
      <Values xsi:type="IntegerVectorType">
        <Vector>1 1 0 0 0</Vector>
      </Values>
    </AdaptationOperator>
    <AdaptationOperator iOPinRef="TEMPORALLEVELS">
      <Values xsi:type="IntegerVectorType">
        <Vector>1 0 1 0 0</Vector>
      </Values>
    </AdaptationOperator>
    <AdaptationOperator iOPinRef="QUALITYREDUCTION">
      <Values xsi:type="FloatVectorType">
        <Vector>1.00 1.00 0.97 0.92 0</Vector>
      </Values>
    </AdaptationOperator>
    <UtilityRank>1 2 3 4 5</UtilityRank>
  </Module>
  <IOPin id="BANDWIDTH"/>
  <IOPin id="SPATIALLAYERS"/>
  <IOPin id="TEMPORALLEVELS"/>
  <IOPin id="QUALITYREDUCTION"/>
</Description>
```

Fig.7 AdaptationQoS description of Football video

As shown in the AdaptationQoS description, the bitrate of switching points are respectively 400 kbps, 220 kbps, 104 kbps, 76 kbps.

Now suppose the connection bitrate is 1000 kbps, using the method of Section 3.3, the decision engine can determine the values of operators: *SpatialLayers*= 0, *TemporalLevels*=0, *QualityReduction*=0.64 (on segment O_1P_1). The visual comparison of the original content and the adapted one is shown in Fig.8. As expected, we can see that the adapted video has the same spatio-temporal resolution as the original one, but that the visual clarity is much blurred.

(a)

(b)

Fig.8 Visual comparison of (a) original video at 2.4 Mbps and (b) adapted video at 1 Mbps (CIF@30fps)

If the bitrate connection is just 170 kbps, similarly the decision engine can determine the values of operators: *SpatialLayers*=1, *TemporalLevels*=0, *QualityReduction*=0.97 (on segment O_3P_3). A frame capture of the adapted video is shown in Fig.9.

To check the optimality of the adaptation system, we created some other adapted video versions of different spatio-temporal resolutions at 1000 kbps and 170 kbps bitrates (and some other bitrates as well). Perceptual verification showed that the adapted content guided by the adaptation trajectory is always better than other bitrate-equivalent versions. Interested readers may find decoded video sequences (in YUV format) of these bitstreams at http://hemera.icu.ac.kr/download.html.

Fig.9 A frame capture of adapted video at 170 kbps (QCIF@30fps)

Discussion

The adaptation trajectory based on user preference can be used as prior knowledge to increase the coding and adaptation efficiency. For example, the adaptation trajectory of text & graphics genre suggests that the SVC bitstream may not necessarily include QCIF spatial layer. It can be encoded with only one CIF spatial layer which results in better coding efficiency than coding with multiple spatial layers. On the other hand, due to the nature of the SVC bitstream structure, some spatio-temporal resolutions may not be extracted from the original bitstream. Adaptation trajectory can tell which spatio-temporal operations should be supported in the original bitstream. For example, adaptation trajectory of crowd genre suggests that the lower spatial layer QCIF should be encoded at 30 fps but not 15 fps.

A similar issue is the decision on values of operators that can be dependent on future adaptation to the current output bitstream. Suppose that we know in advance that the output bitstream at current adaptation node will be further adapted at a future node, and that some spatio-temporal resolution(s) are requested at that future node. Then the operator values at the current node should be properly decided so that the requested spatio-temporal resolution(s) are included in the output bitstream.

Even in some extreme cases, at an adaptation node, the requested spatio-temporal resolution may not be available from the input bitstream. For example the input bitstream contains two layers, QCIF@15fps and CIF@30fps, and the requested output bitstream should be QCIF@30fps. In this case some SVC transcoding method is definitely needed.

In the current experiment, we only employ one short video clip. For a long video stream with different scenes, it can be divided into small shots which are treated as separate clips. In this case, each shot will be provided with one AdaptationQoS description.

CONCLUSION

In this work we have studied adaptation of SVC bitstream in the context of MPEG-21 multimedia framework. For interfacing SVC bitstream with MPEG-21 based adaptation system, we proposed three SVC specific adaptation operators and showed how the concept of adaptation trajectory could be represented by MPEG-21 DIA AdaptationQoS description, and then how the operator values could be computed from that representation. For the actual adaptation at bitstream level, we proposed an adaptation procedure to remove the unnecessary NAL units from an SVC bitstream. The results of this study will help enable QoS management of SVC video streaming in an efficient and standardized manner.

For future work, we will focus on more efficient adaptation procedure, especially on the truncation of FGS and CGS layers. Also, the functionality of ROI scalability in SVC bitstream (Thang *et al*., 2005a) will be integrated in our adaptation system.

References

Chang, S.F., Vetro, A., 2005. Video adaptation: concepts, technologies, and open issues. *Proceedings of the IEEE*, **93**(1):148-158. [doi:10.1109/JPROC.2004.839600]

ISO/IEC IS 15938-5:2001, 2003. Information Technology—Multimedia Content Description Interface—Multimedia Description Schemes.

Joint Scalable Video Model (JSVM) 3.0 ITU-T VCEQ JVT-P202, 2005. Poznan, Poland.

Jung, Y.J., Ro, Y.M., 2004. Joint control for hybrid transcoding using multidimensional rate distortion modeling. Proc. ICIP2004. Singapore.

Jung, Y.J., Thang, T.C., Lee, J., Ro, Y.M., 2003. Visual Media Adaptation System for Active Media. Proc. 2003 Int. Conf. on Imaging Sci., Syst., and Tech..

Jung, Y.J., Kim, Y.S., Kim, D.Y., Kim, J.G., Hong, J.W., Ro, Y.M., 2005a. Analysis of Human Perception for Semantic Concept-based Video Transcoding. Int. Workshop Advanced Image Tech. (IWAIT2005), **1**:251-256.

Jung, Y.J., Thang, T.C., Ro, Y.M., 2005b. Distortion Measures in MPEG-Compressed Domain for Multidimensional Transcoding. Proc. IEEE MMSP2005. Shanghai.

Mukherjee, D., Delfosse, E., Kim, J.G., Wang, Y., 2005. Optimal adaptation decision-taking for terminal and network quality-of-service. *IEEE Trans. Multimedia*, **7**(3):454-462. [doi:10.1109/TMM.2005.846798]

Schwarz, H., Marpe, D., Wiegand, T., 2004. SNR-Sclable Extension of H.264/AVC. Proc. ICIP 2004. Singapore.

Thang, T.C., Bae, T.M., Jung, Y.J., Ro, Y.M., Kang, J.W., Choi, H., Kim, J.G., 2005a. SVC CE8: Spatial Scalability of Multiple ROIs. ITU-T VCEG JVT-O042. Poznan, Poland.

Thang, T.C., Jung, Y.J., Ro, Y.M., 2005b. Effective adaptation of multimedia documents with modality conversion. *EURASIP Signal Processing: Image Communication Journal*, **20**(5):413-434. [doi:10.1016/j.image.2005.03.004]

Thang, T.C., Kim, Y.S., Kang, J.W., Ro, Y.M., Kim, J.G., 2005c. SVC Video Adaptation with MPEG-21 DIA AdaptationQoS. ISO/IEC JTC1/SC29/WG11 m12638. Nice, France.

Vetro, A., 2004. MPEG-21 Digital item adaptation: enabling universal multimedia access. *IEEE Multimedia*, **11**(1): 84-87. [doi:10.1109/MMUL.2004.1261111]

Wang, Y.K., Thang, T.C., 2005. Revised Signaling of Scalability Information. ITU-T VCEQ JVT-P113. Poznan.