MPEG-21-Based Scalable Bitstream Adaptation using Medium Grain Scalability

Hosik Sohn, Hana Yoo, Young Bok Lee, Cheon Seog Kim, Wesley De Neve, Yong Man Ro*

***Abstract*— H.264/AVC Scalable Video Coding (SVC) aims at providing high video quality to users in heterogeneous multimedia usage environments. In order to create an adapted bitstream in an optimal way, an adaptation decision-taking algorithm is needed. This algorithm has to take into account the capabilities and the constraints of the targeted usage environment, as well as the properties of the scalable bitstream. In this paper, description tools part of MPEG-21 Digital Item Adaptation (DIA) are used to characterize usage environments and scalable bitstreams, such as Usage Environment Description (UED), Universal Constraints Description (UCD), and Adaptation Quality of Service (AdaptationQoS). Further, we also propose an adaptation decision-taking algorithm for the exploitation of Medium Grain Scalability (MGS) in SVC. This algorithm was implemented in an Adaptation Decision-Taking Engine (ADTE). Our experimental results demonstrate the efficiency of the proposed adaptation decision-taking algorithm for the exploitation of MGS in SVC.**

***Index Terms*—ADTE, MGS, MPEG-21 DIA, SVC**

I. Introduction

Quality of Service (QoS) is a major goal when creating a Universal Multimedia Access (UMA) environment [1]. To guarantee QoS in diverse usage environments, video adaptation is an essential technique. H.264/AVC Scalable Video Coding (SVC) [2] is the most recent example of a One-Source Multi-Use (OSMU) based video coding method: it allows creating compressed video bitstreams that can be adapted in multiple ways to meet the capabilities and constraints of diverse usage environments. Further, an Adaptation Decision-Taking Engine (ADTE) can be used to determine a feasible adaptation strategy, taking into account the scalability properties of a particular bitstream and the capabilities and constraints of a particular usage environment [3], [4], [5].

In order to communicate the properties of a usage environment to an ADTE, description tools part of MPEG-21 Digital Item Adaptation (DIA) can be used [6], such as Usage Environment Description (UED), Universal Constraints Description (UCD), and Adaptation Quality of Service (AdaptationQoS). The use of these description tools has been described before in [7], [8], and [9].

SVC supports three forms of scalability: spatial, temporal, and Signal-to-Noise Ratio (SNR) scalability. Spatial scalability is realized by coding the video data at multiple spatial resolutions, where information from lower spatial layers is reused to code information at higher spatial layers. Temporal scalability is achieved through the use of hierarchical B pictures [10], while SNR scalability consists of three tools: Coarse Grain Scalability (CGS), Medium Grain Scalability (MGS), and Fine Grain Scalability (FGS) [11], [12]. Compared to FGS and CGS, MGS can provide a feasible number of extraction points at a relatively low complexity. Note that FGS is not incorporated in the current edition of the SVC standard, due to its high complexity.

In this research, we propose an adaptation decision-taking algorithm for exploiting scalability in SVC-compliant bitstreams, putting the focus on the use of MGS. In order to investigate the efficiency of the proposed adaptation decision-taking algorithm, we have constructed a test bed that contains an MPEG-21-based ADTE. Our experimental results show that the proposed adaptation decision-taking algorithm is able to create video bitstreams for diverse usage environments with a feasible quality.

This paper is organized as follows: Section II discusses a number of background technologies that help to understand the proposed adaptation decision-taking algorithm. Section III describes the proposed adaptation decision-taking algorithm in more detail. Experimental results are presented in Section IV, while Section V concludes this paper.

II. Background Technologies

This section outlines a number of technologies that help to understand our proposed adaptation decision-taking algorithm.

A. MGS

Initially, SNR scalability was only made possible by FGS and CGS. FGS allows exercising a fine-grained control of the extraction points at the cost of a high complexity, offering a high number of extraction points. As for CGS, the number of extraction points is significantly lower compared to FGS, while coming with a complexity that remains relatively low.

MGS utilizes a similar coding method as CGS. In particular, MGS can be seen as an enhanced version of CGS, introducing modified high-level signaling and the concept of key pictures. Modified high level signaling enables switching between different MGS layers in any access unit. The concept of key pictures enables an appropriate tradeoff between drift and enhancement layer coding efficiency for hierarchical prediction structures. In short, by making use of MGS, it is possible to

The authors are with the Image and Video Systems Lab, Information and Communications University (ICU), Daejeon 305-732, Republic of Korea (e-mail: yro@icu.ac.kr).

achieve an increased flexibility in terms of bit rate adaptation at the cost of a limited complexity.

B. Description of the Usage Environment

For describing the properties of usage environments and scalable bitstreams, the UCD, UED, and AdaptationQoS description tools can be used. UED describes a usage environment in terms of terminal and network characteristics, natural environment properties, and user preferences [6]. For instance, UED makes it possible to express that the device of an end-user is a laptop, having support for SVC and connected to a network with an average bandwidth of 500 Kbps. UCD allows specifying additional constraints on top of a usage environment description (see further). Finally, AdaptationQoS provides detailed information on the way a scalable bitstream can be adapted, and how this influences the QoS [14].

```
- <Description xsi:type="AdaptationQoSType">
  - <Module xsi:type="UtilityFunctionType">
    - <Constraint iOPinRef="Nominal Bitrate">
      - <Values xsi:type="IntegerVectorType">
          <Vector>1510 1359 1260 1200 1200 1071 1061 1050 1020 941</Vector>
        </Values>
      </Constraint>
    - <AdaptationOperator iOPinRef="Spatial Layers">
      - <Values xsi:type="IntegerVectorType">
          <Vector>0 0 0 0 0 1 1 1 1 1</Vector>
        </Values>
      </AdaptationOperator>
    - <AdaptationOperator iOPinRef="Temporal Levels">
      - <Values xsi:type="IntegerVectorType">
          <Vector>0 1 1 2 3 0 1 2 2 3</Vector>
        </Values>
      </AdaptationOperator>
    - <AdaptationOperator iOPinRef="Priority Id">
      - <Values xsi:type="IntegerVectorType">
          <Vector>63 56 48 40 0 24 16 8 4 0</Vector>
        </Values>
      </AdaptationOperator>
```

Fig. 1. A simplified AdaptationQoS description.

Fig. 1 shows an example of a simplified AdaptationQoS description. Bit rate, spatial scalability, temporal scalability, and priority id are described for each extraction point. As for spatial and temporal scalability, the scalability values described in the actual NAL unit header can be obtained by subtracting the scalability values described in the AdaptationQoS description from the maximum scalability values used in the bitstream. Therefore, contrary to the scalability values described in a NAL unit header, the scalability values described in AdaptationQoS represent a higher quality (higher resolution, higher frame rate) as they approach 0. Further, it is important to know that the perceptual video quality increases as the value of priority id increases.

Fig. 2 shows an example of a simplified UCD description, constraining the spatial resolution, the frame rate, and the value of priority id. In particular, the device of a particular end-user is restricted to CIF resolution, a frame rate of 30 Hz, and a maximum value for priority id of 40. Due to space limitations, other constraints are omitted in Fig. 2.

C. Adaptation decision-taking for SVC

As shown in Fig. 3, bitstreams compliant with SVC can be adapted in various ways in order to take into account different network conditions, device resolutions, and supported frame rates. Fig. 3 demonstrates the use of an ADTE. An ADTE is capable of determining an optimal adaptation method [8], using information stored in the different MPEG-21 descriptions about the characteristics of a usage environment and the properties of a scalable bitstream.

```
- <dia:LimitConstraint>
    <!--  MEI:18 vertical resolution  -->
    <dia:Argument xsi:type="dia:SemanticalRefType" semantics=":MEI:18" />
  - <dia:Argument xsi:type="dia:ConstantDataType">
    - <dia:Constant xsi:type="dia:FloatType">
        <dia:Value>288.000000</dia:Value>
      </dia:Constant>
    </dia:Argument>
    <dia:Operation operator=":SFO:38" />
    <!--  SFO:38  bool( y<=x )  -->
  </dia:LimitConstraint>
- <dia:LimitConstraint>
    <!--  MEI:20 frame rate  -->
    <dia:Argument xsi:type="dia:SemanticalRefType" semantics=":MEI:20" />
  - <dia:Argument xsi:type="dia:ConstantDataType">
    - <dia:Constant xsi:type="dia:FloatType">
        <dia:Value>30.000000</dia:Value>
      </dia:Constant>
    </dia:Argument>
    <dia:Operation operator=":SFO:38" />
    <!--  SFO:38  bool( y<=x )  -->
  </dia:LimitConstraint>
  <!--  AQoS:1.3.9.5  Priority id  -->
- <dia:LimitConstraint>
    <dia:Argument xsi:type="dia:SemanticalRefType" semantics=":AQoS:1.3.9.5" />
  - <dia:Argument xsi:type="dia:ConstantDataType">
    - <dia:Constant xsi:type="dia:IntegerType">
        <dia:Value>40</dia:Value>
      </dia:Constant>
    </dia:Argument>
    <dia:Operation operator=":SFO:38" />
    <!--  SFO:38  bool( y<=x )  -->
  </dia:LimitConstraint>
```

Fig. 2. A simplified UCD description.

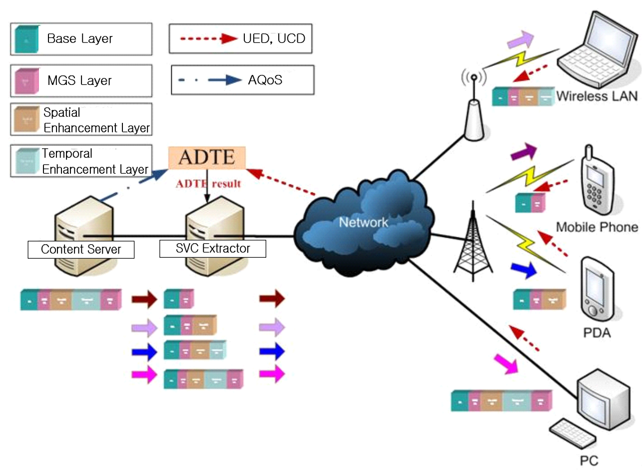

Fig. 3. Scenario illustrating the adaptation and consumption of SVC bitstreams.

Our ADTE consists of an XML parser and an adaptation decision-taking algorithm, as illustrated in Fig. 4. The XML parser extracts information about the usage environment and the scalable bitstream from the MPEG-21 descriptors [6]. Using this information, an adaptation decision is taken, which is then signaled to the extractor. The decision takes the form of scalability coordinates [*D*, *T*, *P*], where *D*, *T*, and *P* denote spatial scalability, temporal scalability, and priority id, respectively. Finally, the extractor creates an adapted bitstream, taking the scalability coordinates and the original bitstream as an input. Since the ADTE uses information about the usage environment of each end-user, personalized consumption of multimedia content can be made possible.

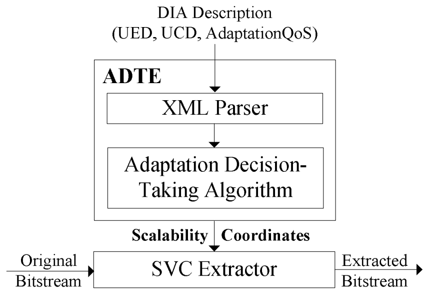

Fig. 4. Internal modules of our ADTE.

III. ADAPTATION DECISION-TAKING

A. *Proposed Adaptation Decision-Taking Algorithm*

In order to extract a bitstream that is suited for a particular usage environment, an ADTE requires an algorithm that determines a number of feasible extraction points based on a description of the usage environment. Fig. 5 represents a flow chart for our proposed adaptation decision-taking algorithm. In Fig. 5, D', T', and P' respectively represent spatial scalability, temporal scalability, and priority id, as communicated in the AdaptationQoS description. Further, D'_{req}, T'_{req}, and P'_{req} denote the minimum spatial scalability, temporal scalability, and maximum priority id described in the UCD description, while T'_{max} denotes the minimum temporal scalability.

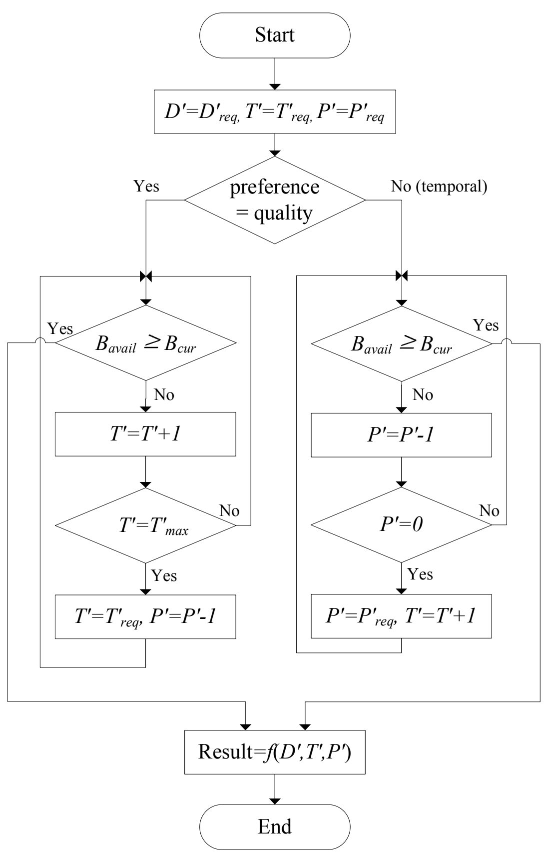

Fig. 5. Flow chart for the adaptation decision-taking algorithm.

As explained in the previous section, the scalability values signaled in the AdaptationQoS description are the opposite of the values communicated in a NAL unit header. Finally, B_{avail} represents the available network bandwidth, while B_{cur} represents the bit rate at the current extraction point with scalability coordinates D', T', and P'.

The proposed adaptation decision-taking algorithm takes into account the spatial resolution, the frame rate, and the visual quality. Our algorithm first gives priority to the spatial resolution, and subsequently allows selecting perceptual quality over temporal resolution, or vice versa, dependent on the preferences of the end-user. Since the Human Visual System (HVS) is the most sensitive to spatial scalability, among the three conventional types of scalability, the highest priority is given to preserving the spatial resolution [15]. In addition, the bandwidth of the network, i.e. the bit rate per second, is also used as an input. In Fig. 5, D', T', and P' are set to the quality values as described in the UCD.

The end-user may prefer a higher perceptual quality over a high frame rate, or vice versa. Let us assume that the end-user gives a higher priority to the perceptual quality. In a next step, the bit rate at the initial scalability coordinates D', T', and P' is compared to the available network bandwidth. If the $Bitrate_{cur}$ is higher than the $Bitrate_{avail}$, the quality requested by the end-user cannot be supported under the current network conditions. In that case, the frame rate is decreased till the bit rate of the bitstream meets the given network bandwidth. If this constraint cannot be met, the picture quality is decreased as well. Finally, using the AdaptationQoS description, the values obtained for D', T', and P' are converted to appropriate NAL unit header values. The converted values f(D', T', P') represent the final decision taken by our ADTE.

B. *ADTE Processes*

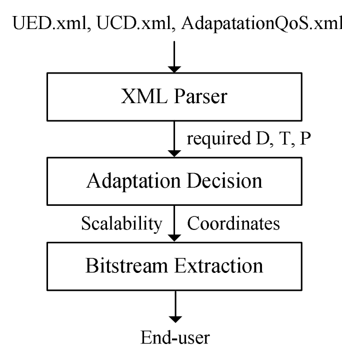

Fig. 6. Processes within our ADTE.

Fig. 6 outlines the different processes within our ADTE. First, initial values for the spatial resolution, frame rate, priority id, and the network bandwidth are computed by analyzing the UED and UCD descriptions. After comparing the acquired information with the information in the AdaptationQoS description, the final values for spatial scalability, temporal scalability, and priority id are determined using the adaptation decision-taking algorithm discussed in the previous section. The extractor then adapts the original bitstream, based on the scalability coordinates received from the ADTE. At this point in time, the extractor decides whether a NAL unit should be kept or not by checking the scalability coordinates in the header of all NAL units of the original bitstream. Finally, the extracted bitstream is transmitted to the end-user and the terminal decodes the received bitstream.

The proposed system utilizing an ADTE for the adaptation of scalable bitstreams is depicted in Fig. 7. The server contains an ADTE for determining the feasible extraction points. The server also contains an adaptation module, which is responsible for the actual bitstream extraction.

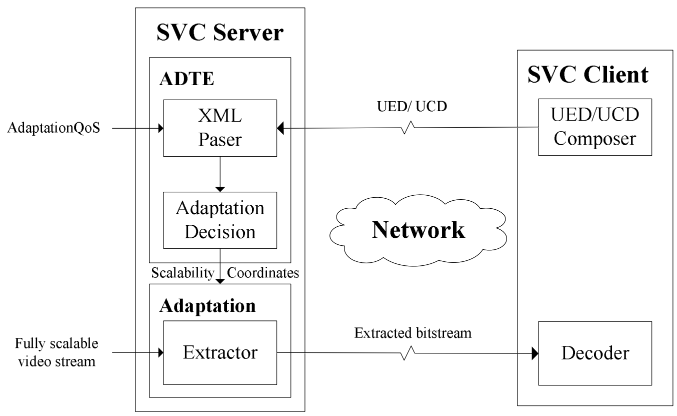

Fig. 7. Overview of the proposed system.

IV. Experiments

In order to verify the usefulness of the proposed adaptation decision-taking algorithm, we compared the use of MGS to the use of CGS. For our experiments, we encoded the Football test sequence with JSVM 8.9 [16], using the settings in Table I.

TABLE I
ENCODING PARAMETERS.

Spatial	Temporal	QP
QCIF	3.75, 7.5, 15	40, 34
CIF	3.75, 7.5, 15, 30fps	43, 37

In a first experiment, we compared the use of extraction points in CGS and MGS, assuming that an end-user wants CIF resolution video at a frame rate of 30 fps, while having to meet a network bandwidth of 900 kbps. Furthermore, the end-user wants to have a minimal perceptual quality, by giving priority id a value of 40.

The red line in Fig. 8 represents the network bandwidth. The bit rates obtained at the different extraction points are also visualized in Fig. 8. The X-axis in the graph indicates the different extraction points as offered by MGS and CGS, while the Y-axis indicates the accumulated bit rates. In CGS, extraction points are determined by dependency id, temporal level, and quality level. On the other hand, for MGS, extraction points are determined by dependency id, temporal level, and priority id.

As shown in Fig. 8, CGS only supports one extraction point that is below the network bandwidth constraint of 900 kbps. In particular, CGS is able to produce a video bitstream with CIF resolution, a frame rate of 30fps, and a bit rate of 477.66 kbps. On the other hand, MGS supports three extraction points when taking into account the network bandwidth constraint of 900 kbps. In particular, for MGS, the system produces a video bitstream with CIF resolution, a frame rate of 30fps, and a bit rate of 893.44 Kbps. The corresponding PSNR values also prove that MGS with 29.27 dB provides a better video quality than CGS with 25.44 dB. In CGS, the flexibility in terms of bit rate adaptation is restricted due to a limited number of available extraction points. However, for MGS, the number of extraction points is higher, offering more flexibility in terms of bit rate adaptation.

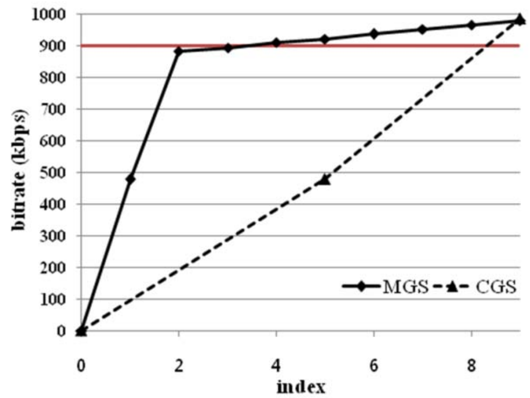

Fig. 8. Bit rates according to the available extraction points.

Fig. 9 shows the bit rate of the extracted bitstream according to a varying network bandwidth. More precisely, the bandwidth is gradually increasing from 470 Kbps to 1000 Kbps in a non-monotonous way.

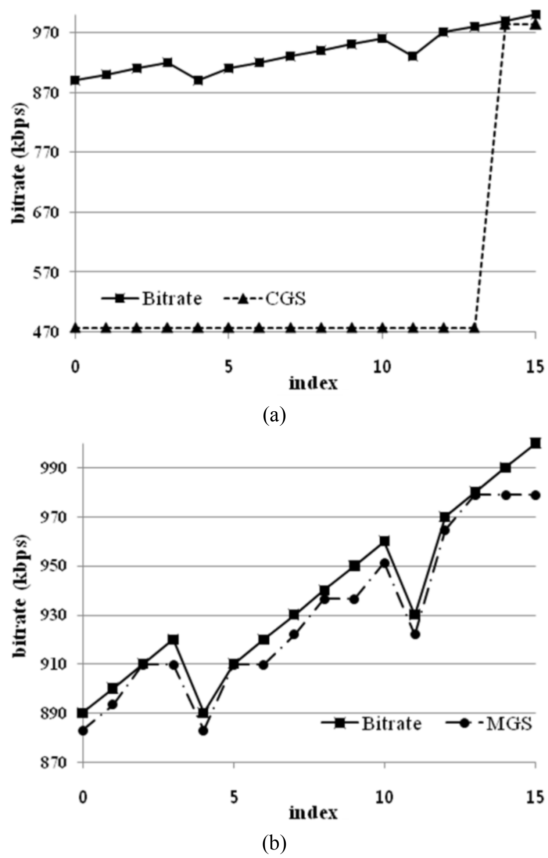

Fig. 9. Bit rate of extracted bitstreams: (a) CGS and (b) MGS.

In Fig. 9, the X-axis in the graph indicates the MGS and CGS extraction points, while the Y-axis indicates the obtained bit rates. In order to better differentiate between the graph of the network bandwidth and the experimental results, the range of the Y-axis in Fig. 9(a) is different from the range of the Y-axis in Fig. 9(b).

Comparing both results illustrates that MGS is better suited to deal with network bandwidth variations. Indeed, as for CGS, the extracted bitstreams are not able to fully use the available network bandwidth. Since the amount of bits used by an extracted bitstream is directly related to the perceptual quality of a picture, MGS is typically able to provide users with better video quality than CGS for a given network bandwidth.

Overall, our experimental results show that our proposed adaptation decision-taking algorithm, in combination with MGS, allows creating adapted bitstreams with a high quality, and where the bitstreams are suited for use in diverse usage environments.

V. CONCLUSION

In this paper, we proposed an adaptation decision-taking algorithm that is suited for the adaptation of scalable video using MGS. In order to demonstrate the usefulness of the adaptation-decision taking algorithm, we constructed a test bed containing an MPEG-21-based ADTE. The capabilities and constraints of the usage environment are described by MPEG-21 DIA description tools, as well as the properties of the scalable bitstreams. Information about both the usage environment and the scalable bitstream that is to be adapted is taken into account during the adaptation decision-taking.

The usefulness of the adaptation decision-taking algorithm is verified by two methods: the first method compares the extraction points in CGS and MGS, while the second method compares the bit rate of the extracted bitstreams with respect to a varying network bandwidth. The experimental results show that the proposed decision-taking algorithm is able to efficiently adapt bitstreams in order to take into account a varying network bandwidth. Since MGS is a coding technique that is able to solve the shortcomings of FGS and CGS, the proposed adaptation decision-taking algorithm can be efficiently used to adapt SVC video bitstreams, coded with MGS and thus offering diverse extraction points, while the encoding complexity remains relatively low.

REFERENCES

[1] A. Perkis, Y. Abdeljaoued, C. Christopoulos, T. Ebrahimi, and J. F. Chicaro, “Universal multimedia access from wired and wireless systems,” Birkhauser Boston Transactions on Circuits, Systems, and Signal Processing, Vol. 20, No. 3-4, pp.387-402, May, 2001.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, Sep., 2007.

[3] J. W. Kang, S. H. Jung, J. G. Kim, and J. W. Hong, “Development of QoS-aware Ubiquitous Content Access (UCA) Testbed,” *IEEE Transactions on Consumer Electronics*, Vol.53, Issue 1, pp. 197 - 203, Feb., 2007.

[4] D. Mukherjee, E. Delfosse, J. G. Kim, and Y. Wang, “Optimal Adaptation Decision-Taking for Terminal and Network Quality-of-Service,” *IEEE Transaction on Multimedia*, Vol. 7, Issue 3, pp. 454 - 462, June., 2005.

[5] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Real-Time System for Adaptive Video Streaming Based on SVC," *IEEE Transaction on Circuits and Systems for Video Technology*, Vol. 17, Issue 9, pp.1227-1237, Sept. 2007.

[6] DIA ISO/IEC 21000-7:2004, Information Technology – Multimedia Framework – Part 7: Digital Item Adaptation, 2004.

[7] I. Kofler, C. Timmerer, and H. Hellwagner, “Efficient MPEG-21 based adaptation decision-taking for scalable multimedia content,” *Proc. Of the 14th Multimedia Cumputing and Networking Conference (MMCN’07)*, Jan/Feb, 2007.

[8] T. C. Thang, Y. S. Kim, Y. M. Ro, J. W. Kang, and J. G. Kim, “SVC bitstream adaptation in MPEG-21 multimedia framework,” *15th International Packet Video Workshop (PV2006)*, vol. 7, no, 5, pp.764-772, May, 2006.

[9] T. C. Thang, J. W. Kang, Y. S. Kim, Y. M. Ro, and J. G. Kim, “SVC Video Adaptation with MPEG-21 DIA AdaptationQoS,” *ISO/IEC JTC1/SC29/WG11 M12638*, Oct., 2005

[10] H. Schwarz, D. Marpe, and T. Wiegand, “Analysis of hierarchical B-pictures and MCTF,” *IEEE International Conference on Multimedia & Expo (ICME)*, pp. 1929-1932, Jul., 2006.

[11] L. Amonou, N.Cammas, S.Kervadec, and S.Pateux, “Enhanced SNR scalability for layered CGS coding using quality layers,” *Joint Video Team JVT-S044*, Apr., 2007.

[12] M. Wien, H. Schwarz, and T. Oelbaum, “Performance Analysis of SVC,” *IEEE transaction on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp.1194-1203, Sep., 2007.

[13] T. C. Thang, J. W. Kang, Y. S. Kim, Y. M. Ro, J. G. Kim, S. Pateux, and S. Devillers, "CE report on SVC and AdaptationQoS," ISO/IEC JTC1/SC29/WG11 M12915, Jan., 2006

[14] J.Magalhães and F. Pereira, “Using MPEG standards for multimedia customization,” *Signal Processing Image Communication*, vol. 19, no. 5, pp. 437-456, May, 2004.

[15] H. Knoche, J. D. McCarthy, M. A. Sassse, “Can small be beautiful? assessing image resolution requirements for mobile TV,” *in Proc. The 13th annual ACM multimedia*, pp. 829-838, November, 2005.

[16] “Joint Scalable Video Model JSVM – 8,” *JVT-U202*, Oct., 2006