Scalability adaptation for optimisation of coding efficiency of SVC-based video

H. Kodikara Arachchi[1], C.T.E.R. Hewage[1], S. Dogan[1], M. Mrak[1],
V. Barbosa[2], M.T. Andrade[2], and A.M. Kondoz[1]
[1] I-Lab, CCSR, University of Surrey, Guildford, GU2 7XH, Surrey, UK
[2] INESC Porto, Rua Dr. Roberto Frias 378, 4200-465 Porto, Portugal
E-mail: *h.kodikaraarachchi@surrey.ac.uk*

Abstract - *The explosion of interest in consuming rich multimedia content resulted in a growing demand for technologies that enable seamless access over any network using any terminal without affecting the viewing experience. Scalable Video Coding is considered to be one of the key technologies that will improve the accessibility to video resources through adaptation at network nodes at a negligible computational cost. State-of-the-art video coding technologies enable a comprehensive set of scalability options. However, the trade-off of having such enriched set of scalability options is the reduced compression efficiency and increased encoder/decoder complexity. This trade-off can be mitigated if the encoder has prior knowledge of the scalability options and levels that are actually needed at a given time. It is possible to achieve this goal in a closely monitored system, such as a Virtual Collaboration System (VCS), in which users and their usage environment are known. This paper proposes a framework to determine the scalability options and levels for encoding scalable video for VCS. Simulation results highlight the benefits of the proposed framework.*

Keywords – *Video Adaptation, Scalable Video Coding, Virtual Collaboration System, H.264/AVC*

1. INTRODUCTION

Today's multimedia communication landscape has been greatly shaped by a number of coexisting complementary as well as competing codec, access, delivery and consumption technologies. With this heterogeneity of the underlying technologies in mind, guaranteeing the Quality of Experience (QoE) [1][2] expected by users is a nontrivial exercise. In addition to the technological factors, the diversity of user preferences, as well as when and where the content is consumed add additional dimensions to the already complex dilemma. As a result, under the umbrella of the Universal Multimedia Access (UMA) concept, the notion of transparent access to rich multimedia content is widely discussed in the research community [3]. The MPEG-21 standard, one of the most recognised efforts to pave the way to the success of UMA, has promoted the content adaptation to achieve the goals of the UMA [4].

Scalable Video Coding (SVC) is one of the key technologies that will improve the accessibility to video resources through adapting them at network nodes or the destination at a negligible computational cost. Furthermore, it enhances the resilience to network deficiencies that may significantly disturb the QoE. State-of-the-art video coding standards offer a comprehensive set of scalability options, including spatial, temporal and Signal-to-Noise Ratio (SNR). Granularity of each scalability option is extended by defining scalability levels. For instance, a bit stream may have three spatial scalability levels, corresponding to QCIF, CIF and 4CIF resolutions. However, the trade-off of having such enriched set of scalabilities is the reduced compression efficiency and increased encoder/decoder complexity. This trade-off can be mitigated if the encoder has prior knowledge of the scalability options and levels that are actually needed for the users and transport network. Based on this information, the encoder can produce a bit stream having only the required scalability options and levels. It is possible to achieve this goal in a closed system, such as a Virtual Collaboration System (VCS) [5], in which users and their usage environment are known. This paper discusses a framework for adapting the scalability options for encoding scalable video for such a VCS application. Under the proposed framework, an Adaptation Decision Engine (ADE) deduces the scalability structure for each encoder based on the usage context. Simulation results presented in the paper highlight the benefits of the proposed framework.

2. SCALABLE VIDEO CODING

Compressed video bit-streams that are produced using scalable coding can be adapted using specialised low-complexity techniques which involve bit stream parsing only. Adaptation points are related to video representation with combination of different frame sizes (spatial scalability), different frame rates (temporal scalability) and available bit-rates (quality scalability). With recent technology developments the scalable video coding also achieves high compression and enables wide range of combination of different scalabilities [6]. New

SVC standard is, besides capabilities to provide combination of temporal, spatial, and quality scalability, able to support base layer compatibility with H.264/AVC.

In scalable video coding standard, the temporal scalability is supported by hierarchical temporal prediction structures. In these structures, key pictures are coded at regular intervals by using only previous key pictures as references. The pictures between the key pictures are the hierarchical B pictures, which are bi-directionally predicted from the key pictures. The base layer contains a sequence of the key pictures at the coarsest supported temporal resolution, while the enhancement layers consist of the hierarchically coded B pictures. In such system a low-delay coding can be realised by restricting the prediction of the enhancement layer pictures from only previous frames.

Spatial scalability is achieved by using a multi layer coding approach. It relies on motion compensation at each required spatial level and inter-layer prediction. Inter-layer motion, residual or intra prediction mechanisms are used to improve the coding efficiency of the enhancement layers.

The scalable extension of H.264/AVC supports quality scalability using coarse-grain scalability (CGS) and medium-grain scalability (MGS). CGS is achieved using spatial scalability concepts with the exclusion of the corresponding up-sampling operations in the inter-layer prediction mechanisms. MGS is introduced to improve the flexibility of bit-stream adaptation and error robustness. Furthermore, this improves the coding efficiency of the bit-streams which aim at providing different bit rates.

3. ADAPTATION OF SCALABILITY

The content adaptation framework proposed for the VCS consists of four core components: (1) ADE, (2) Adaptation Authoriser (AA), (3) Context Providers (CxP), and (4) Adaptation Engines (AE) [7]. Even though all those components of the platform are equally important for the functionality of a successful content adaptation system, in this paper, we mainly focus on that of the ADE and AE, since the operations of other two components are not altered.

In general, an ADE evaluates the current usage environment constraints and conditions as informed by one or more CxPs and determines the adequate adaptation operations to maximise the user experience. AEs are the functional blocks that perform the adaptation operations as advised by the ADE. Both AEs and ADE can be located in any terminal within the network, including the source and the destination nodes, depending on the requirement. For example, an operation that improves the QoE of multiple users can be performed either at the source node or an intermediate nodes while operations that addresses personal level constraints, such as displaying subtitles, are performed at the user terminal.

When a virtual collaboration session is considered, the required scalability options and levels that are useful at a given time instance depends on a number of factors. Some of the leading factors are described below:

(*a*) *Diversity of terminal capabilities*: Decoding of higher resolution layers may not be advantageous if the spatial and/or temporal resolution of the terminal display is much lower than that of the video. In such a scenario, a resolution layer matches with that of the terminal device would enable the system to forward the relevant information only. On the other hand, insufficient computational resources may force the user terminals to discard some information without decoding. For example, if the device has resources just enough for decoding a QCIF resolution video and if the smallest resolution provided in the bit stream is CIF, the terminal device then cannot decode, and hence cannot display the media at all. Therefore, it is clear that having lower resolution scalability layers could improve the user satisfaction under such circumstances.

(*b*) *Network condition*: During a congestion period, for example, an intelligent network node can prioritise scalability layers based on their relative importance. Hence, adding an extra low resolution/SNR scalability layer enables the router dropping the enhancement information only. As a result, the desirable options and levels of scalability depend on the severity of the congestion.

(*c*) *Different user preferences*: Even if the terminal and network capabilities allow going for a higher resolution video, the user may prefer to view the video on a smaller window on his/her personal computer while working on the desktop in the background in the mean time.

Since the VCS is a closely monitored system, in which the participants are known at the time when they are connected to the system, an ADE can easily determine the scalability structure so that the overall system performance is maximised. The AE for this adaptation is a real-time encoder which resides within each of the virtual collaboration terminal.

3.1. Adaptation decision

In a typical multi-party virtual collaboration session, a number of terminals of various calibres interact with each other using various access networks with vast range of capabilities. One or more scalability options may be used by the AEs located within the communication networks and on the terminal devices to address the prevailing usage environment conditions. The proposed adaptation decision model, which is illustrated in the Fig. 1, determines the useful adaptation options and levels required for serving these usage conditions as appropriate.

To deduce the required scalability options, the ADE needs to collect and interpret contextual information. To achieve this, the ADE keeps track of the instantaneous adaptation options and levels required by each participant as well as the network availability during the virtual collaboration session. This contextual information is realised by the ADE as constraints imposed by the delivery and consumption environment as well as the capabilities of the video encoders. The ADE implements an optimisation algorithm, by which it selects the set of service parameters for a particular encoder so as to maximise a given utility, while satisfying the context constraints. The utility can be some objective measure of the user satisfaction, such as the quality of the audiovisual stream. This relationship can be expressed using the MPEG-21 Digital Item Adaptation (DIA) Adaptation Quality of Service (AQoS) tool [8][9]. The usage context conditions can be expressed using the MPEG-21 DIA Universal Environment Description (UED) tool. The constraints can be derived from the low-level contextual information, and higher level concepts can then be inferred using a specific ontology developed for use in this particular application scenario. Fig. 2 provides a high-level view of the ADE architecture that can be employed to take the necessary adaptation decisions.

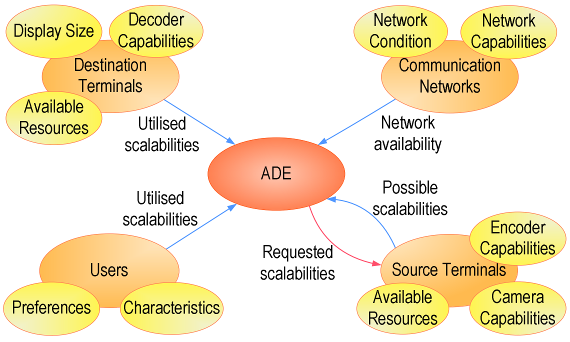

Fig. 1. Adaptation of scalability

The developed ontology provides a formal definition of the concepts and their relationships as well as of the domain of the application, *i.e.*, the virtual collaboration application. The ADE uses the sensed context to create instances of the concepts and relations defined by the ontology, thus obtaining a better knowledge of the prevailing real-world situation. This information can then be used to assist the adaptation decision process. Fig. 3 illustrates the flow diagram of the complete decision process.

3.2. The encoder

As mentioned above, the encoder acts as the AE for the adaptation discussed in this paper. The proposed encoder is designed to be capable of changing the scalability structure on the fly based on the instructions received from the ADE. In our experiments, we used the emerging scalability extension of H.264/AVC [6] as the video encoding technology.

One of the possible ways of changing the scalability structure is to use an IDR frame [10]. When the decoder decodes an IDR frame, it clears the reference frame buffer [10]. This operation marks the start of a new prediction chain, which effectively ceases the further referrals to the frames decoded before the IDR frame. Therefore, if it is used in conjunction with the necessary Sequence Parameter Sets (SPS) and Picture Parameter Sets (PPS), it is possible to start a new scalability structure within the same video stream. The proposed algorithm is illustrated in Fig. 4.

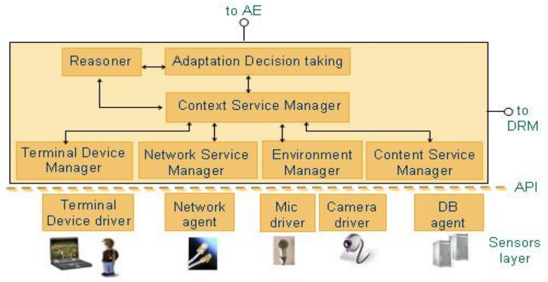

Fig. 2. High level architecture of the ADE

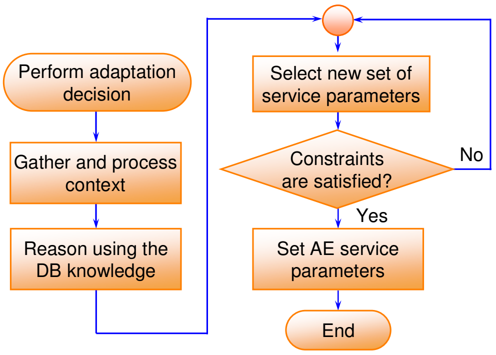

Fig. 3. The adaptation decision process

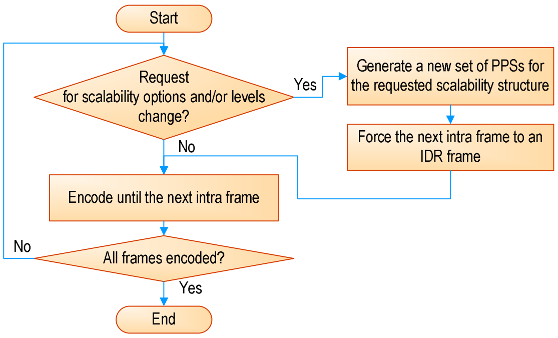

Fig. 4. The proposed algorithm for changing the scalability structure of the scalable H.264/AVC

4. SIMULATION RESULTS

The simulation results presented in this section highlight the effectiveness of adapting the number of scalability layers for the given usage requirement. In the tests conducted, the JSVM 9.4 is used as the software platform. Fig. 5 and Fig. 6 present a comparison of the objective quality variations of bit rate adaptation for the ‘Mobcal’ and ‘Parkrun’ video test sequences, respectively. Two scalable source bit streams are generated for each test sequence and they are used as the original encoded video sequences, from which other spatial, temporal and quality (bit rate) levels are selected. Scalable bit streams are obtained by varying the number of spatial scalability levels. The highest spatial and temporal resolutions are 1280×704 pixels and 50 frames per second (fps) (progressive), respectively. Lower spatio-temporal resolutions are dyadic subdivisions of the maximum resolution.

The rate-distortion performances of the highest temporal and spatial scalability layers are illustrated in Fig. 5 for the two test sequences in view. From this result, it can be noted that over a 0.5 dB rate-distortion improvement can be obtained by limiting the number of scalability layers. This performance loss is due to the inclusion of an extra low resolution spatial scalability layer and two MGS layers. Moreover, when a lower resolution layer is selected from the aforementioned video streams, the rate-distortion gain of limiting the number of spatial resolution layers reaches as high as 2.5 dB, as shown in Fig. 6. This clearly shows that managing the scalability levels adequately improves the rate-distortion performance during content adaptation.

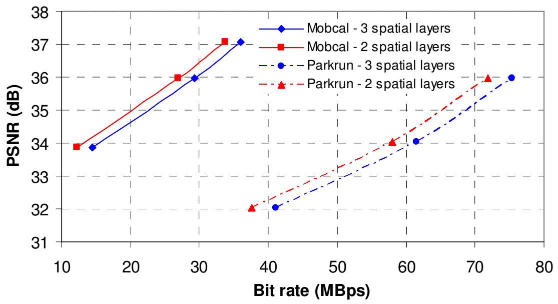

Fig. 5. Results for decoding at original spatial and temporal resolution (1280×704 at 50 fps)

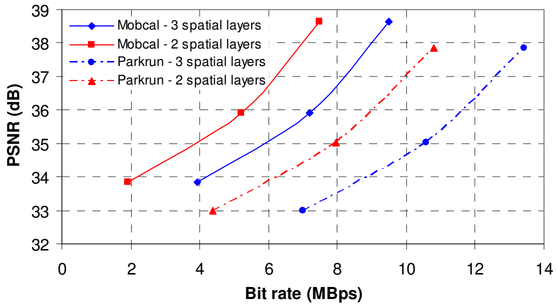

Fig. 6. Results for decoding a lower spatial and temporal resolution (640×352 at 25 fps) from the original bit stream

4. CONCLUSION

A context-aware content adaptation framework for adapting the scalability options and levels for scalable video coding for virtual collaboration applications is discussed. An ADE obtains information on the utilised scalability options and deduces the scalability structure for the encoder. The simulation results highlight the importance of such adaptation and demonstrate that as high as 2 dB objective quality gains can be achieved through adapting the scalability options and levels for the application scenario.

ACKNOWLEDGEMENT

The work presented was developed within VISNET II, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme.

REFERENCES

[1] F. Pereira, “A triple user characterization model for video adaptation and quality of experience evaluation”, in *Proc. 7th IEEE Multimedia Sig. Process. Workshop*, Oct. 2005, pp. 1-4.

[2] P. Reichl, “From ‘quality-of-service’ and ‘quality-of-design’ to ‘quality-of-experience’: A holistic view on future interactive telecommunication services”, in *Proc. 15th Int. Conf. Software, Telecommun. and Computer Networks*, Sep. 2007, pp. 1-6.

[3] F. Pereira and I. Burnett, “Universal Multimedia Experiences for Tomorrow”, *IEEE Sig. Process. Mag.*, vol. 20, no. 2, Mar. 2003, pp. 63-73.

[4] I. Burnett, R Van de Walle, K. Hill, Bormans, J. F. Pereira, “MPEG-21: Goals and Achievements,” *IEEE Multimedia*, vol. 10, no. 6, pp. 60-70, Oct.–Dec. 2003.

[5] C.T.E.R. Hewage, H. Kodikara Arachchi, T. Masterton, A.C. Yu, H. Uzuner, S. Dogan, and A.M. Kondoz, “Content adaptation for virtual office environment using scalable video coding”, in *Proc. 16th IST Mobile and Wireless Commun. Summit* Jul. 2007.

[6] 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, Sep. 2007, pp. 1103-1120.

[7] V. Barbosa, A. Carreras, H. Kodikara Arachchi, S. Dogan, M. T. Andrade, J. Delgado, and A. M. Kondoz, “A scalable platform for context-aware and DRM-enabled adaptation of multimedia content”, accepted to *ICT Mobile and Wireless Communications Summit 2008*, June 2008.

[8] *Information Technology – Multimedia Framework (MPEG-21) – Part 7: Digital Item Adaptation*, ISO/IEC Standard, ISO/IEC 21000-7:2004, Oct. 2004.

[9] I. Burnett, F. Pereira, R. Walle, and R. Koenen *eds*., “*The MPEG-21 Book*”, John Wiley and Sons Ltd, 2006.

[10] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Recommendation H.264, March 2005.