A Quality-of-Experience Video Adaptor for Serving Scalable Video Applications

Chia-Ho Pan, I-Hsien Lee[1], Sheng-Chieh Huang[2], Chung-Jr Lian, and Liang-Gee Chen, *Fellow*, IEEE

Abstract — *When transmitting video over a heterogeneous network, it is required to satisfy the different constraints due to the preferences and equipment selections of different users. Scalable video coding is one of the solutions to serve this problem. More than one video parameters include spatial frame size, temporal frame rate, and visual quality resolution are utilized to provide better scalability in scalable video coding scheme. It is difficult to find the relationship between the various video parameter settings and user preferences. In this paper, we propose a quality-of-experience video adaptor applied on scalable video applications. This adaptor adjusts video parameters for each user appropriately and achieves a better visual quality by using negligible computation. The simulated results show the accuracy of adaptation points can match the user's preference from 78% to 93% by using proposed scheme. This solution provides not only a practical way in video adaptor design, but also an efficient scheme to achieve better subjective view.*

Index Terms — Scalable video coding, objectivity-derived modeling, video adaptor, multidimensional adaptation selection.

I. INTRODUCTION

Flexibility or scalability of video stream, which are not emphasized in conventional video coding, significantly affects system performance of video applications in heterogeneous networks. For this reason, scalable video coding techniques are widely used among present networks to deal with various requirements of clients. In Fig. 1, different clients may want to decode a single bitstream with a great diversity of user preferences, such as network conditions, equipment constraints, power consumption and so on. Obviously, conventional video coding technique with a rigid configuration cannot satisfy such a requirement and a more flexible design is preferable in this user-oriented era. Several video parameters which include spatial frame size, temporal frame rate, and visual quality resolution can be utilized to provide a better scalability and visual quality under different user requirements. Once the video parameters requested by clients are different from those supported by content providers, video adaptors [1] will appropriately adjust these parameters to fit certain requirement. Thus, it is believed that video adaptor plays a dominating role in the area of scalable video coding.

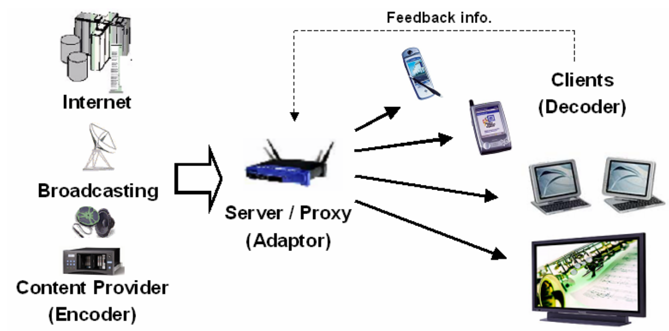


Fig. 1. Role of adaptor in scalable video coding application.

Prior works for video adaptation [2]-[4] focus on transcoding efficiency. These techniques convert the original video bitstream from a specific compression format or video parameter setting to another without encoding the video sequence again. A video parameter setting is a combination of video parameters which reflect the specific video parameters based to the different preferences and device limitations of clients. In other words, transcoding related works transfer specific adjustments according to individual requests of clients one by one. In this case, a large amount of users simultaneously having different requirements implies a corresponding high computation demand in transcoding techniques. This potential problem lets transcoding techniques not suitable to serve multiple users. Another technique which is capable of transmitting video bitstream over heterogeneous network is called embedded bitstream techniques [5]-[10]. These techniques embed the properties of scalable videos in the encoded video bitstream. By applying embedded bitstream techniques, one video bitstream can serve various requirements from clients without format conversion. In this paper, the proposed approach also bases on embedded bitstream technique so as to provide more appropriate video adaptation and reduce possible computation in transmitting scalable video bitstream over heterogeneous network.

For embedded scalable video coding, MPEG-4 **Error! Reference source not found.** provides a well-known coding tool named fine granular scalability (FGS). By using this tool, users can have the feature of visual quality scalability which is

This work was supported in part by the TW. Department of Commerce under Grant No. 95-221-E-002-195, and by ASUSTek Computer Inc..

The authors are with the DSP/IC Design Lab, Graduate Institute of Electronics Engineering and Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan (Email: lgchen@video.ee.ntu.edu.tw).

[1]I-H. Lee is with the Department of Electrical Engineering, Pennsylvania State University, PA, USA (Email: izl107@psu.edu)

[2]S.-C. Huang is with the Department of Electrical and Control Engineering, National Chiao-Tung University, HsinChu, Taiwan (Email: schuang@cn.nctu.edu.tw)

Contributed Paper
Manuscript received June 30, 2007

also called signal to noise ratio (SNR) scalability. MPEG-4 FGS [5] delivers layered video data and uses only one parameter, quantization parameter (QP), to control the relationship between bitrate and distortion. QP is a simple but useful index to find optimal solution in the single dimensional case of SNR scalability.

In addition to SNR scalability, spatial and temporal scalabilities are also included in many scalable video techniques [7]-[10]. Scalable Video Coding (SVC) [10] supports spatial, temporal, and SNR scalabilities in the same video stream. By using SVC, users can obtain the possibility of selecting different spatial frame size, temporal frame rate, and visual quality resolution. In these coding techniques, more than one video parameter is embedded in scalable video stream. As a consequence, in multi-dimensional scalability, it is complex to choose an appropriate video parameter setting to fulfill the requirement from clients. Since multidimensional selection is essentially a problem of multidimensional video adaptation, a systematic approach to multidimensional video adaptation [12] was developed to transform the content objects of a multimedia document, where modality conversion is applied to achieve content adaptation. The classification-based adaptation scheme [13] using a multidimensional adaptation prediction method that combines the temporal and SNR scalability by using subjective quality evaluation scheme. Although the computation of previous work is less than video decoder, the heavy computation is still an problem of serving various requirements from numerous clients in the networks at the same time.

In this paper, we address the multidimensional video adaptation and purpose an objectivity-derived scheme to facilitate the selection of multidimensional adaptation. Moreover, the proposed video adaptor identifies key frames in the video sequences and re-allocates their bitrates to improve visual quality. This adaptor will appropriately adjust video parameters for each user to achieve a better visual quality by using negligible computation. In other words, this solution provides not only a practical way in video adaptor design, but also an efficient scheme to achieve better subjective view.

This paper is organized as follows. In Section II, we present the concept of quality-of-experience video adaptor. Section III describes the characteristics of our proposed multidimensional video adaptor. Section IV shows the experimental results. Finally, a conclusion is given in Section V.

II. Quality-of-Experience Video Adaptor

Video adaptor plays an important role in the applications of scalable video framework. User preferences or equipments' constraints may introduce different requirements of spatial frame size, temporal frame rate, SNR quality resolution, and so on. Thus, different video parameters need to be determined in according to the given requirement. However, it is complex to construct analytical models in theoretical optimization to obtain the relationship between the video parameters and the subjective impression of each user. When the spatial-temporal resolution is fixed, parameter adaptation problem of single dimensional SNR scalability becomes the case QP determination as described in FGS. While spatial-temporal resolutions can also vary with client's request, the video adaptation is more complicated than single dimension case. Consequently, in multidimensional video adaptation, we will need a more precise index to evaluate the performance of extracted bitstream combination.

In this paper, we propose a quality-of-experience video adaptor for serving scalable video applications. The model of multidimensional adaptation decision is based on the objective emulation tool such as peak signal to noise ratio (PSNR). PSNR is commonly used for measuring picture quality degradation of visual impression. For simplicity, PSNR is a widely-used scheme for objective evaluation. However, conventional quality measure schemes such as PSNR and mean square error (MSE) may not properly represent the users' perspective of video quality (e.g., [14]-[16]). The users' perspective of video quality is relative to the trend of PSNR curve in single dimensional case, but the detail of this relation is not derived in multidimensional case. In this paper, we develop the model of multidimensional adaptation decision based on the concept of objectivity-derived evaluation scheme. The derived PSNR-like curve is utilized to emulate the SNR scalability in multidimensional adaptation first. For the temporal and spatial scalabilities, we extrapolate different PSNR-like curves in each spatial-temporal resolution to emulate the subjective preference of each user. Then the objectivity-derived scheme can be utilized as the selection model for multidimensional adaptation appropriately for the video adaptation between different spatial-temporal resolutions. Moreover, the preference of each user is also important for video adaptor to make adaptation decision. We design a statistical learning scheme to obtain the subjective preference of each user's experience according to feedbacks when selecting the video parameters. By doing this, a better selection of preferred visual quality in multidimensional adaptation for each individual user can be yielded according to the using experiences from each user.

After selecting a preferred video parameter setting in multidimensional adaptation, the proposed video adaptor identifies the key frames in the video sequences to improve the visual quality. Some frames may not have much influence on visual quality when the available bits of the frames are changed. In our design, the key frames in a group of picture (GOP) are identified by video adaptor to achieve better visual quality by adjusting available bits of the frames according to the video context and the hierarchy of temporal scalability. By doing this, we utilize the network bandwidth more efficiently under the same network bandwidth for the application of heterogeneous network.

Many applications of video streaming such as IPTV use the automatic mode to select the appropriate video parameter of scalable video stream automatically. The prior works selects the adapted video stream by using a predefined model which

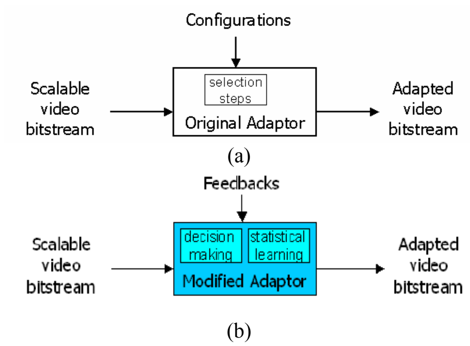


Fig. 2. Block diagram of (a) original adaptor engine, (b) modified adaptor engine.

is not specified by the user. The original video adaptor engine of SVC, which takes in two inputs: a scalable video stream and configurations as shown in Fig. 2(a). This adaptor extracts the adapted video stream according the configurations decided in advance. The configurations provide the decisions of video parameter settings, and specify how to adapt a scalable video stream. Since this scheme only uses a generic model for all users, it can not take user's using experience to further improve the performance of future adaptation. As we have mentioned, the preference of each user is important for video adaptor in making a adapting decision. Fig. 2(b) is the block diagram of modified adaptor engine. The modified video adaptor engine of SVC takes two inputs: the scalable video stream and the feedbacks. The feedbacks such as available network bandwidth and user's video parameter setting of the current information of the clients are reported to the server in order to get a more suitable adapted video stream for clients. The proposed adaptor engine, based on these inputs, yields the adapted scalable video stream a better quality of experience for each individual user. Briefly speaking, we use the feedback from individual user to be the input of reference learning and key frame identification, to select appropriate modeling which will be applied when selecting the video parameters. After these adjustments, an appropriate video parameter setting which fits the subjective preference of each user is selected, and this combination can achieve a better subjective performance than the original video adaptor. To demonstrate the effect between our design and original video adaptor, we apply these designs on standard SVC codec [10].

III. Characteristics of Proposed Multidimensional Video Adaptation

In this section, we introduce the characteristics of the proposed objectivity-derived video adaptor as follows.

A. Objectivity-Derived Model for Multidimensional Selection

Choosing a parameter setting is the essential part of multidimensional parameter adaptation. For single dimensional scalability, PSNR curve is a popular objective tools for video quality measurement because it provides correlation to the subjective impression. When the resource is

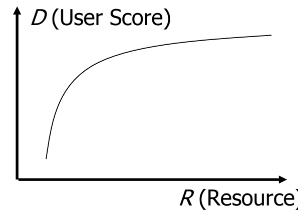


Fig. 3. Rate-Distortion concept.

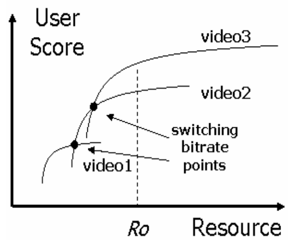


Fig. 4. Emulation for multidimensional scalable video coding

increasing, the user score increases and then saturates according to the user point of view. In other words, both PSNR and user score will increase, then, saturate when the available resource is increased as shown in Fig. 3. The resources mentioned here may include limited network bandwidth, computation complexities, available power consumption, user's equipment constrain and so on. To simplify the problem, in this paper, we only consider the limited network bandwidth in the following discussion. Other possible constraints, which may affect the parameter selection, can be considered in a similar manner. In our design, the user score is modeled according to the above description by the following equation

$$\textit{User socre} \;\; \alpha \;\; \log(x) \tag{1}$$

where x is the scalability percentage of available bitrate. In saturated area, the user score is proportional to x gradually. For simplicity, we model the user score is linear proportional to x. Hence, the model of the SNR scalability in multidimensional adaptation is emulated by the PSNR-like curve as shown below.

$$z = a + b * \log(x) + c * y, 0 < x \le 100, y = x/100 \tag{2}$$

In this equation, z is the user score, a is the initial point of user score for current spatial-temporal resolutions, b and c are the scaling factor to adjust the emulated curve and y is the scalability ratio varied from 0 to 1 about the truncated bitrate ratio of SNR scalability.

After objectivity-derived model is established, it is simple to make decision on parameter selection so as to achieve SNR scalability. However, it is not easy to find the best match to user's preferences if several parameters are all adjustable at the same time. Many possible choices can be choosed at the same resource constraint. Therefore, it is necessary to have a model which includes multiple parameters in making the decision. If the model of multidimensional adaptation exists or can be built, it is easy to choose the right video parameter setting that maximizes the user score. On the contrary, it is difficult to derive an impartial emulation tool for evaluating the relationship of user score between spatial and temporal scalabilities. The user score is the subjective preference that varies from user to user. For different combinations of spatial and temporal scalabilities, it is not easy to represent different

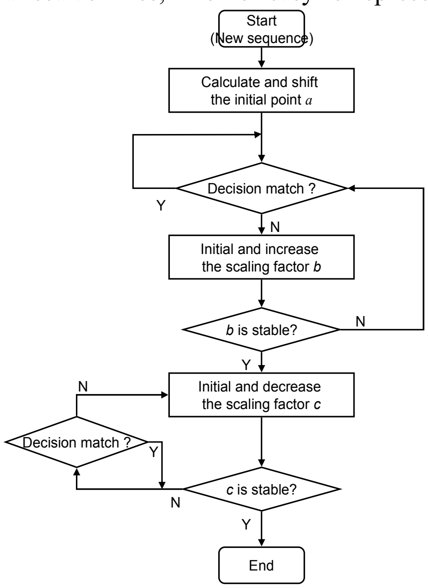


Fig. 5. The flow chart of the statistical learning scheme.

sensory feeling of each user among the different video parameter setting.

To solve this problem, the basic idea of our scheme is to extend the trend of objective evaluation tool, the PSNR-like curve, as the multidimensional selection model. For SNR scalability, we use the PSNR-like curve to model the dimension of SNR scalability. For spatial and temporal scalability, we take different independent curves to model the complex relationship between different spatial-temporal resolutions. The different PSNR-like curves represent the SNR scalability of corresponding spatial-temporal resolution. Fig. 4 shows three video streams with different spatial-temporal resolutions. For evaluating the user score more precisely, not only single curve is used to represent all different spatial-temporal resolutions. Compared this with other schemes, our method does not have to consider the complex intuitions of each user and impartial evaluations of different spatial-temporal resolutions in one evaluation curve or model. Different PSNR-like curves provide more flexibilities of user score between different spatial-temporal resolutions to build the model for multidimensional adaptation selection. Note that the temporal, spatial, and SNR scalabilities can be modeled simply and separately by different PSNR-like curves.

Comparing with a generic model for all users, the user-dependent model is more appropriate and flexible because different preferences and equipment selections from individual users are considered. Therefore, a statistical learning scheme is designed to reflect the user subjective preference of spatial and temporal scalabilities. We shift and scale the PSNR-like curves to match the feedbacks of the video parameters from each user. The different RD-like curves cross the switching bitrate points as shown in Fig. 4. The distinct switching bitrate points are the bitrate which the preferred video parameter setting changes. These points are useful to provide the prediction of the adaptation operation

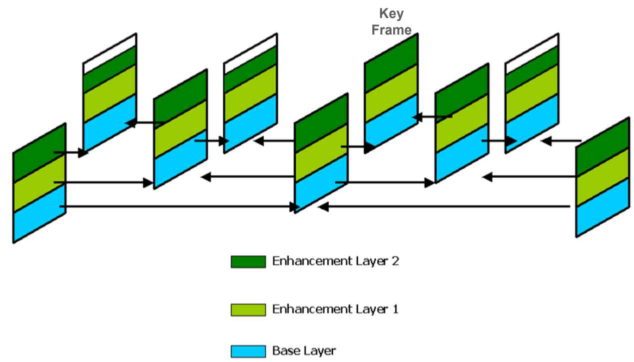


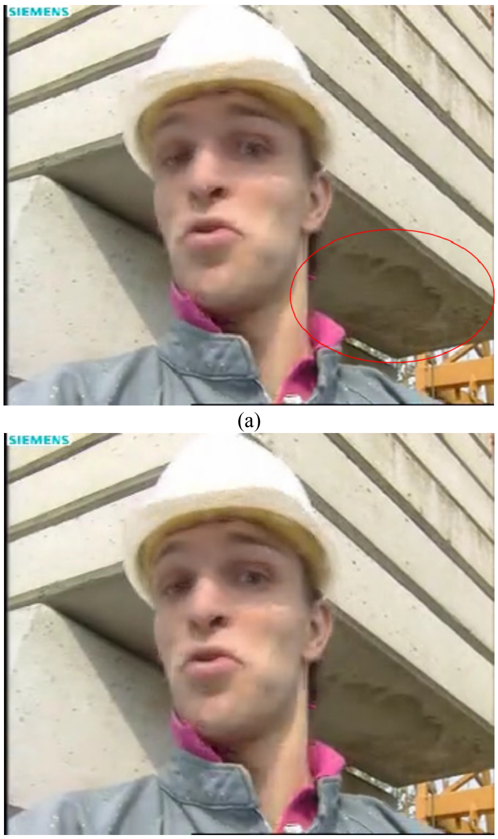
Fig. 6. Temporal hierarchical relation in a GOP.

with different spatial and temporal resolution at any given bandwidth. Fig. 5 is the flow chart of the statistical learning scheme. The initial point *a* should be decided first because it represents the relationship of user score in different spatial-temporal resolution of each user. For different spatial scalability, the initial point *a* of higher spatial scalability is determined by the average of the available emulated user scores with the same bitrate but smaller spatial size. After deciding the rough user score location of each spatial-temporal resolution, scaling factor *b* and *c* are used to learn the user experience for refining the scale and the tendency in saturation region of objectivity-derived model. With the user-dependent model, the proposed objectivity-derived scheme can provide more appropriate adjustments of video parameters of each user.

B. Efficient Bandwidth Utilization

Bandwidth is limited and variant for the users in heterogeneous network applications. After selecting a video parameter setting in multidimensional adaptation, utilizing the limited bandwidth more efficiently to improve the visual quality is also important.

The temporal scalability of SVC is processed by motion compensated temporal filtering (MCTF). According to the temporal process relationship, the lower temporal level frames are the basic reference frames in a GOP. In other words, the lower temporal level frames are referenced several times to provide information for other different higher temporal level frames. The natural of inter-frame referencing in video coding leads to a potential problem of error propagation once the frame to be referenced is corrupted. If the lower temporal level frames have quantization error or mismatch error, the drift error propagates to those frames which reference this frame. For this reason, we modify the bit allocation scheme of the frame in each temporal level. We allocate higher amount of bits for the lowest frame first in order to reduce the drift error between encoder and decoder. The pixels in lower temporal level frames are closer to the accurate pixel value when it is used as the reference frame for the frames in other temporal levels. Thus, the degradation of video quality can be significantly reduced. It reduces the drift error caused by the prediction mismatch between reference frame and processed frame as shown in Fig. 6. This concept also can be used in another temporal scalability scheme, hierarchical B frame architecture, which is applied in H.264/AVC **Error! Reference source not found.**. The following is the pseudo code of the modified process procedure

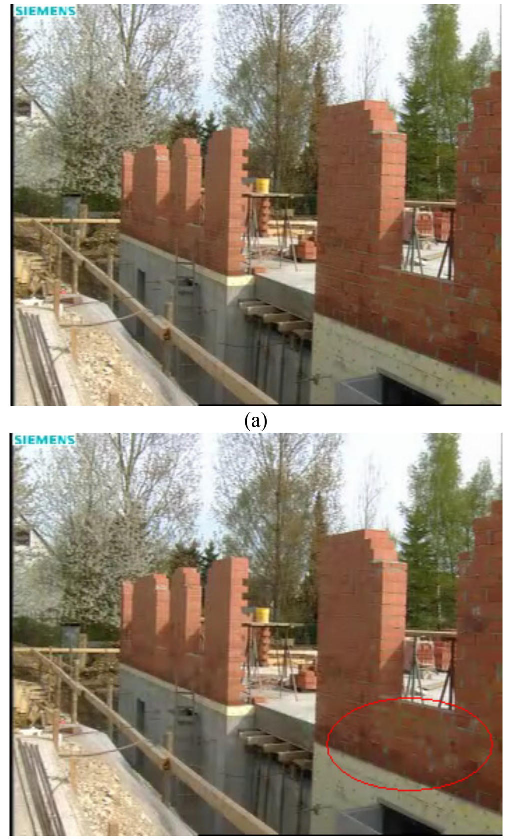


(a) (b)

Fig. 7. Illustrations of temporal hierarchical bit allocation effect at the same constraint(CIF, 260kpbs). (a) frame 87 of modified temporal hierarchical concept of Foreman sequence (b) frame 87 of non-modified temporal hierarchical concept of Foreman sequence.

for(temporal level)
for(spatial layer)
for(FGS layer)
{
bandwidth judgment ;
}

The GOP of SVC is directly proportional to the power of 2 according to the hierarchy of MCTF and hierarchical B frame. Because the GOP size is usually fixed when the video is encoded, it is possible that the key frame is located in any frame of any temporal levels. We identify the key frame according to the bits of each frame in the same temporal level. In this case, we try to allocate more bits for the key frame to improve the visual quality as shown in Fig. 7. The reduced bits of the other frames, caused by the tradeoff of limited remaining bit rate, do not affect the visual quality significantly. On the contrary, the visual quality of key frame is improved by adding more bits on it as shown in Fig 8. We have a better video quality performance than original video adaptor in SVC. Therefore, the bandwidth is used efficiently in the proposed video adaptor.



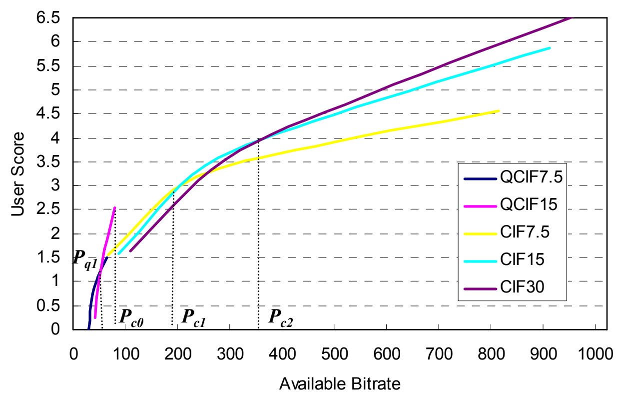
(a) (b)

Fig. 8. Illustrations of key frame effect at the same format(CIF, 260kbps). (a) frame 255 of combined key frame scheme and temporal hierarchical concept in Foreman sequence (b) frame 255 of modified temporal hierarchical concept in Foreman sequence.

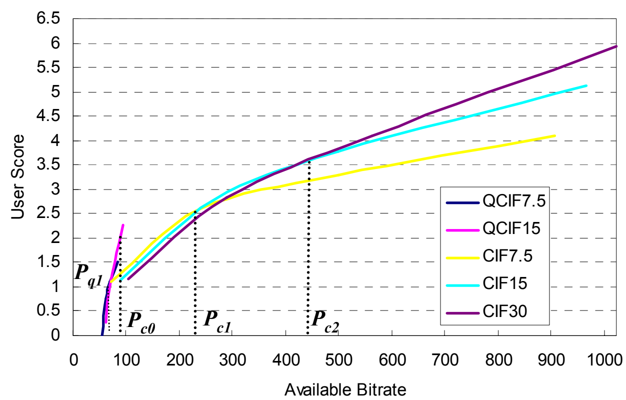
IV. Experimental Results

In this section, we implement our proposed characteristics of multidimensional video adaptation in the video adaptor of SVC reference software [18]. The spatial resolution includes QCIF and CIF size. The QCIF size has 7.5 frames per second (fps) and 15fps for temporal scalability, and the CIF size has 7.5fps, 15fps and 30fps for temporal scalability.

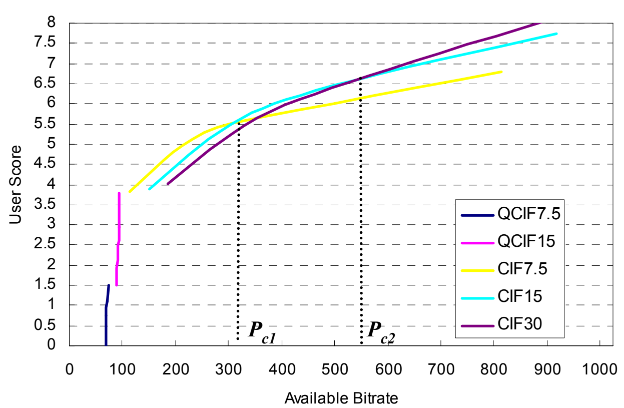
Fig. 9 is the simulation results of multidimensional adaptation selection model of the proposed video adaptor at different bandwidth in different test sequences. From Fig. 9, it is observed from low, medium to high bandwidth, the preferred adaptation of spatial resolution shifts from QCIF to CIF. The preferred temporal frame rate operates from lower frame rate to original frame rate as 7.5 fps, 15 fps to 30fps. The relationship between user score and available network bandwidth varies with the different test sequences. The adaptation points P_{q1}, P_{c0}, P_{c1}, P_{c2} reveal the switching bitrate points to provide the adaptation operation with different spatial and temporal resolution between different temporal frame rate of QCIF and CIF resolutions. From the simulations, the preference of temporal frame rate for this user is not less than 15fps because the preferred video parameter of 15fps between switching bitrate points P_{c1} and P_{c2} is larger than other users. The spatial frame size is more important than temporal frame rate if the larger spatial frame size is available.



(a)



(b)



(c)

Fig. 9. The preferred multidimensional adaptation selection model with overall bandwidth for (a) Foreman sequence, (b) Mobile and Calendar (M&C) sequence, and (c) Stefan sequence.

The adaptation histogram in [13] shows the video adaptation of different frame rates at different bandwidths for subjective simulation. The switching bitrate point of CIF resolution is fixed about 150kbps, 260kbps for Foreman test sequence, 180kbps, 490kbps for M&C test sequence, and 270kbps, 610kbps for stefan test sequence respectively by counting number of subjects' preferring each different adaptation operation at specific bandwidths. However, each user has their preference about the video parameters. A generic model is not appropriate to fit the variable preferences of every user. The proposed method provides a user dependent adaptation model for video adaptation application.

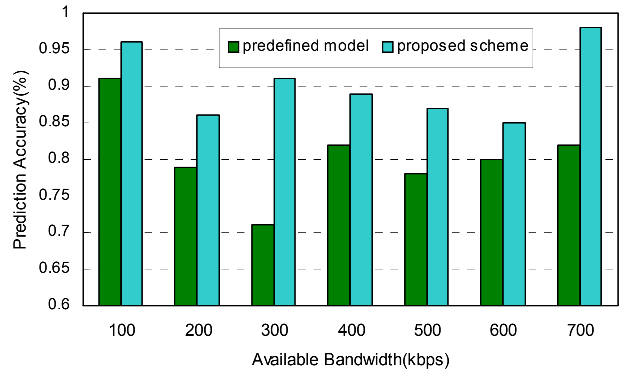


Fig. 10. Prediction accuracy comparison in predefined model and proposed scheme between 100kbps to 700kbps.

Compared with a generic model, the R-D like curves can be adjusted to reflect the user's objective preference when different test sequences. From Fig. 9, the switching point of CIF resolution is adapted to 190kbps, 360kbps for Foreman test sequence, 220kbps, 450kbps for M&C test sequence, and 310kbps, 560kbps for Stefan test sequence respectively according to preferring each different adaptation operation of the current user. In the experimental results, we see that the proposed model is intuitive and reflect the user's preferences adequately.

When the proposed model of video adaptor is stable, the model can be used as current user's selection model for further video sequences. The test sequences can be divided by three corresponding categories with low, medium and high content complexity, respectively. The minimal achievable bandwidth [13] by any possible adaptation of each bitstream is used as an indicator of the video content complexity criteria for video categories. We use the similar simulation environment as [13] for evaluating the perceived impairment of each user. We adopt the single stimulus continuous quality evaluation (SSCQE) recommended by ITU-R standard [19] with some revision to simulate adaptation of spatial and temporal scalability. This choice is made because SSCQE is more suitable to the evaluation of long sequences and it relates well to the time-varying quality of test sequences. The possible adapted test sequences (clips) are aligned in one row. The test clips are generated every 50kbps from minimum adaptation bitrate to maximum adaptation bitrate. The display window of different spatial, temporal, bandwidth combination was randomized. Fig. 10 is the prediction accuracy of different spatial and temporal scalability prediction between 100kbps and 700kbps. The proposed method improves the average accuracy prediction rate from 78% to 93% in overall available adaptation bandwidth of test sequences by counting 18 subjects. For different spatial and temporal scalability, the proposed method can be used for the further video bitstreams for scalable video applications.

After multidimensional adaptation prediction, we improve the visual quality under given bandwidth. The PSNR of the modified temporal hierarchical scheme outperforms the original one. Our proposed video adaptor can identify the key frames in the video sequences and improve the visual quality after selecting a video parameter setting in multidimensional adaptation. Fig. 11 is the frame by frame luminance PSNR comparison between original video adaptor and the proposed video adaptor in SVC. The PSNR value of the proposed scheme outperforms the original one by no more than 1dB, the subjective video quality is better than original ones. When we allocate more bits for the key frame in the video stream, which can reduce the drift error propagated to other frames, the PSNR is improved under 0.3dB than the modified temporal hierarchy scheme by changing the bitrate of each frame slightly. However, the visual quality of key frame is improved further than the frame without using any concept as shown in

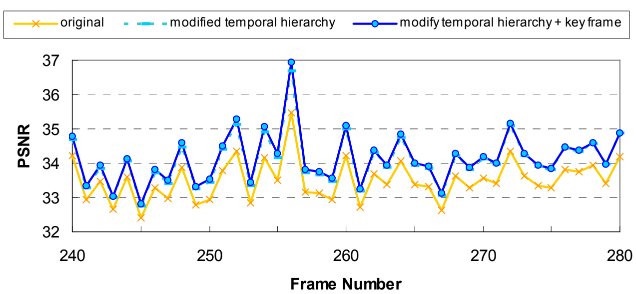


Fig. 11. Frame by frame luminance PSNR comparison between 240 to 280 of Foreman sequence at 224kbps.

Section III. Therefore, we allocate more bits for the really key frames without losing much PSNR in other frames. By combining the temporal hierarchical concept and key frame scheme, we utilize the bandwidth more efficiently and improve the objective and subjective view under the same video parameter setting.

For the computation complexity, Fig. 12 shows the iteration times of initial point a and scaling factors b and c in overall available adaptation bandwidth. The iteration time is different from the preferred bitrate of switching bitrate point. Even for the special case (minimum or maximum available bitrate) of preferred switching bitrate, the iteration times are still under 5 iterations. The overhead computation is negligible. Besides, decoding the compressed scalable video bitstream is not necessary in our proposed quality-of-experience adaptor. Other previous works need to decode or partial decode the compressed bitstream in order to get more information to build the video adaptation model. Hence, it is very simple and useful to fix the user preference rapidly for multidimensional scalable video selection in practical application.

In our proposed quality-of-experience video adaptor, we duplicate the trend of PSNR-like curves to emulate the user score of SNR scalability in different combinations of spatial and temporal scalabilities. We also design a statistical learning scheme to reflect the user subjective preference of spatial scalability and temporal scalability. After the shifting and scaling of each curve by the statistical learning scheme, the PSNR-like curves fit the subjective view of each user. Therefore, it is very simple and useful for multidimensional scalable video selection in practical application over heterogeneous networks.

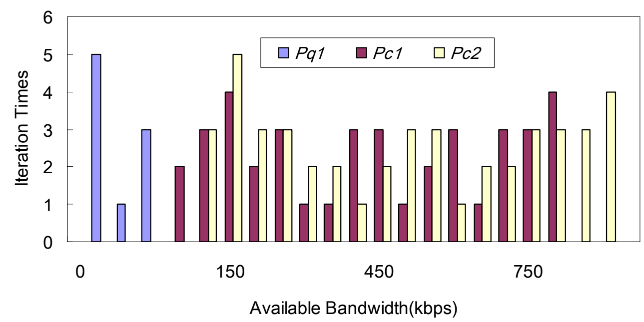


Fig.12 The iteration times of the adaptation points P_{q1}, P_{c1} and P_{c2} comparison for Foreman sequence over appropriate bandwidth.

V. Conclusion

In this paper, we propose a multidimensional adaptation selection scheme to match the preferences of the video parameters for each user. This scheme characterizes the relationship between spatial, temporal and SNR scalabilities according to the subjectiveness of each user. An objectivity-derived emulation scheme is used to realize the selection of multidimensional adaptation. In this scheme, feedbacks from user are provided to strengthen the adaptation selection, and the key frames are identified to enhance the visual quality. This scheme can be used in video proxies or gateways without much computation overhead.

Our proposed video adaptor considers the user feedbacks and the content information at the same time to provide more appropriate adjustments of the video parameters for each user. The proposed method can achieve 93% accuracy prediction rate of when the switching bitrate point is stable. Besides, the video adaptor utilizes the bandwidth in a more efficient way and achieves better subjective visual quality. The experimental results show that the video adaptor provides high consistence of quality between the adjusted video stream and the expectation of users. The proposed scheme is an easy and direct solution in practical design of a video adaptor in related products of scalable video applications.

References

[1] S.-F. Chang and A. Vetro “Video adaptation: concepts, technologies, and open issues,” *Proc. IEEE*, vol. 93, no. 1, pp. 148-158, Jan. 2005.

[2] C.-W. Lin, Y.-C. Chen, and M.-T. Sun, “Dynamic region of interest transcoding for multipoint video conferencing”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 10, pp. 982–992, Oct. 2003.

[3] A. Vetro, C. Christopoulos, C. and H. -F. Sun, "Video transcoding architectures and techniques: an overview", *IEEE Signal Process. Mag.*, vol. 20, no. 2, pp. 18–29, Mar. 2003.
[4] J. Xin, C.-W. Lin, and M.-T. Sun, "Digital video transcoding", *Proc. IEEE*, vol. 93, no. 1, pp. 84-97, Jan. 2005.
[5] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, 2001.
[6] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 332–344, 2001.
[7] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 572–588, 1994.
[8] B. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-DSPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1374–1387, 2000.
[9] P. Chen and J. W. Woods, "Bi-directional MC-EZBC with lifting implementation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 10, pp. 1183-1194, Oct. 2004.
[10] "Scalable Video Coding Working Draft 2," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-O201, Apr. 2005.
[11] "Generic Coding of Audio-Visual. Objects: Part 2 - Visual," ISO/IEC JTC1/SC29/WG11 N1902,. FDIS of ISO/IEC 14496-2 Amd.1, 2000.
[12] T. C. Thang, Y. J. Jung, Y. M. Ro, "Effective adaptation of multimedia documents with modality conversion", *Signal Process.: Image Commun.*, Vol. 20, Issue 5, pp.413-434, 2005
[13] Y. Wang, M. van der Schaar, S. -F. Chang, A. C. Loui, "Classification-based multidimensional adaptation prediction for scalable video coding using subjective quality evaluation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 10, pp. 1270-1279, Oct. 2005.
[14] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, pp. 2959–2965, Dec. 1995.
[15] S. Winkler, "A perceptual distortion metric for digital color video," in *Proc. SPIE*, vol. 3644, 1999, pp. 175–184.
[16] Z.Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[17] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, May 2003.
[18] "JSVM 2.0 Software," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-O203, Apr. 2005.
[19] *Methodology for the subjective assessment of the quality of television pictures*, Recommendation ITU-R BT.500-10, Jun. 2000.

Chia-Ho Pan received the B.S. and M.S. degrees in Electrical Engineering from National Central University, ChungLi, Taiwan, R.O.C., in 1997 and 1999, respectively. He is currently pursuing the Ph.D. degree at the Graduate Institute of Electrical Engineering from National Taiwan University, Taipei, Taiwan, R.O.C..

His major research interests include video coding, video codec VLSI architecture design and wireless multimedia systems.

I-Hsien Lee received the B.S. and M.S. degrees in Electrical Engineering from National Central University, ChungLi, Taiwan, R.O.C., in 1999 and 2001, respectively. He is currently pursuing the Ph.D. degree at the Department of Electrical Engineering, Pennsylvania State University, PA, USA.

Sheng-Chieh Huang was born in ChungHua, Taiwan., in 1967. He received the B.S. degree in Hydraulic Ocean Engineering from Nation Cheng Kung University, in 1991, and the M.S. and Ph.D. degrees in Electrical Engineering from National Taiwan University, in 1993 and 1999.

He is an Assistant Professor with the Department of Electrical and Control Engineering at NCTU. His current research interests are VLSI design in DSP/DIP architecture design, computer architecture, and video coding system.

Chung-Jr Lian received the B.S., M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1997, 1999 and 2003, respectively.

His major research interests include image and video coding (JPEG, JPEG 2000, MPEG, H.264/AVC ...), image and video codec VLSI architecture design. He is now a postdoctoral research fellow at DSP/IC Design Lab, supervised by Prof. Liang-Gee Chen..

Liang-Gee Chen (S'84–M'86–SM'94–F'01) was born in Yun-Lin, Taiwan, R.O.C., in 1956. He received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan, R.O.C., in 1979, 1981, and 1986, respectively.

He was an Instructor (1981–1986), and an Associate Professor (1986–1988) in the Department of Electrical Engineering, National Cheng Kung University. While in the military service during 1987–1988, he was an Associate Professor in the Institute of Resource Management, Defense Management College. In 1988, he joined the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan. During 1993–1994, he was a Visiting Consultant with the Digital Signal Processing (DSP) Research Department, AT&T Bell Labs, Murray Hill, NJ. In 1997, he was a Visiting Scholar with the Department of Electrical Engineering, University of Washington, Seattle. During 2001 to 2004, he was the first Director of the Graduate Institute of Electronics Engineering (GIEE), NTU. During 2004 to 2006, he is also the Director of the Electronics Research and Service Organization, Industrial Technology Research Institute, Hsinchu, Taiwan. Currently, he is a Professor with the Department of Electrical Engineering and GIEE at NTU. His current research interests are DSP architecture design, video processor design, and video coding systems.

Dr. Chen has served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS for Video Technology since 1996, as Associate Editor of IEEE TRANSACTIONS ON VLSI SYSTEMS since 1999, and as Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II since 2000. He has been the Associate Editor of the Journal of Circuits, Systems, and Signal Processing since 1999, and a Guest Editor for the Journal of Video Signal Processing Systems. He is also an Associate Editor of the PROCEEDINGS OF THE IEEE. He was the General Chairman of the 7th VLSI Design/CAD Symposium in 1995 and of the 1999 IEEE Workshop on Signal Processing Systems: Design and Implementation. He is the Past-Chair of Taipei Chapter of IEEE Circuits and Systems (CAS) Society, and is a member of the IEEE CAS Technical Committee of VLSI Systems and Applications, the Technical Committee of Visual Signal Processing and Communications, and the IEEE Signal Processing Technical Committee of Design and Implementation of Signal Processing Systems. He is the Chair-Elect of the IEEE CAS Technical Committee on Multimedia Systems and Applications. During 2001–2002, he served as a Distinguished Lecturer of the IEEE CAS Society. He received the Best Paper Award from the R.O.C. Computer Society in 1990 and 1994. Annually from 1991 to 1999, he received Long-Term (Acer) Paper Awards. In 1992, he received the Best Paper Award of the 1992 Asia-Pacific Conference on circuits and systems in the VLSI design track. In 1993, he received the Annual Paper Award of the Chinese Engineer Society. In 1996 and 2000, he received the Outstanding Research Award from the National Science Council, and in 2000, the Dragon Excellence Award from Acer. He is a member of Phi Tan Phi.