SCALABLE VIDEO ADAPTATION OPTIMIZATION USING SOFT DECISION SCHEME

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

DSP/IC Design Lab, Graduate Institute of Electronics Engineering and
Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[1]TCM/Sense Lab, Electrical and Control Engineering, National Chiao-Tung University,HsinChu,Taiwan
[2]Department of Electrical Engineering, Pennsylvania State University, PA, USA

ABSTRACT

Video transmission over a heterogeneous network suffers a condition that it has required to satisfy many different constraints because of the variety limitations of client equipments and preferences of each user. In this condition, the video stream should have the flexibility feature. Scalable video coding is one of the solutions to serve this condition. In order to provide a better flexibility on scalability feature, more than one combination with video spatial size, temporal frame rate, and visual quality resolution are utilized. It is difficult to adapt the appropriate relation between the various combinations and user preference. In this paper, we propose an objectivity-derived algorithm with soft decision scheme for serving scalable video adaptation. This method can be used in existing equipments without much computation overhead. This useful and simple solution is not only a practical way in scalable video adaptation design, but also an efficient scheme of achieving subjective satisfaction.

1. INTRODUCTION

Scalable video coding is useful for the applications of heterogeneous network to deal with various client's requirements. As shown in Fig. 1, the clients decode the compressed bitstream with a great diversity of user's preferences, network conditions, client equipment constraints, power issues and so on. However, the conventional video coding technique can only support a rigid or one fixed video parameter configuration. Many video parameters which include spatial frame size, temporal frame rate and visual quality resolution are utilized to provide more flexibility to serve various client's requirements. If the video parameters of the decoder are not the same as those of the encoder, the video adaptation technique [1] adapts appropriate video parameters for the decoder according to the different requirements of clients. Thus, the video adaptation plays an important role in the area of scalable video adaptation.

This paper addresses this issue, and develops an objectivity-derived video adaptation scheme to realize the automatic selection of the scalable video adaptation. Many application of video streaming such as IPTV has two modes to adapt the suitable adjusted video stream for the users. The manual mode adapts the video parameters by the selection of each user's preference. The automatic mode selects the appropriate video parameter of scalable video stream automatically by a predefined model. This mode selects the adapted video stream by using a predefined model which is not specified by the user. This paper considers the user's preferences to provide more appropriate adjustments of the adaptation model for each user. In order to strength the adaptation selection, we design a statistical learning scheme of user's preferences to achieve a user dependent selection model. We also design a trellis-based searching scheme to search the suitable bitrate switching point. Then, the proposed scheme provides a user dependent selection model between the adjusted video stream and the preference of users directly in scalable video adaptation.

This work was supported in part by the TW. Department of Commerce under Grant No. NSC95-2221-E-002-195, NSC96-2221-E-002-155, and by ASUSTek Computer Inc.

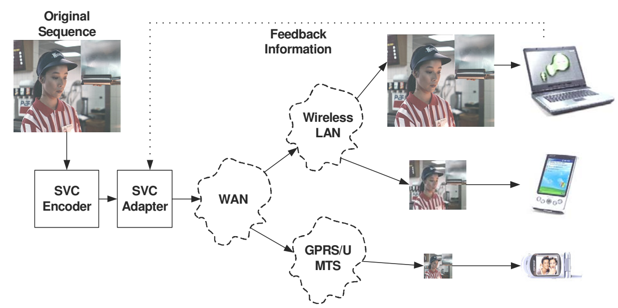

Fig. 1. Scenario of scalable video application.

2. SCALABLE VIDEO ADAPTATION

Prior works for video adaptation [2] focus on efficient transcoding techniques. These techniques convert the original video stream from a specific compression format to another without encoding the video sequence again. When the amount of client is enormous, transcoding techniques can not satisfy various requirements of clients at the same time. Another technique for the applications of heterogeneous networks is called embedded bitstream technique, which embeds the scalability properties in the encoded bitstream. In addition, one encoded bitstream can serve for all requirements of clients at the same time by using embedded bitstream technique. Compared with conventional video techniques, several previous works [3]-[4] acclaim that embedded bitstream technique can provide better flexibility and scalability on video coding.

The MPEG-4 fine granular scalability (FGS) [3], a visual quality scalability also called SNR scalability, is a well-known coding technique that delivers layered video data with precise rate control. MPEG-4 FGS uses only one parameter, quantization parameter (QP), to control the relationship between bitrate and distortion. For single-dimensional case of SNR scalability, it is simple to make a decision to select the QP of single-dimensional scalable video stream.

Besides SNR scalability, spatial and temporal scalabilities are also included in Scalable Video Coding (SVC) [4]. In this coding technique, more flexibility and functionality are provided for user's selection. More than one video parameters need to be determined in scalable video stream compared with MPEG-4 FGS stream relatively. The combination of video parameters named video parameter setting reflects the specific video parameters based to the different preferences and device limitations of clients. It becomes complex to choose an appropriate video parameter setting to match the constraints of a client under multidimensional scalability, multidimensional adaptation selection is essential for multidimensional video adaptation. A systematic approach [5] to multidimensional video adaptation is developed to transform the content objects of a multimedia document, where modality conversion is applied to achieve content adaptation. The classification-based adaptation scheme [6] proposes a multidimensional adaptation prediction method that combines the temporal and SNR scalability by using subjective quality evaluation. The framework [7] proposes the weighted PSNR model for SVC adaptation and this method is used to find the adaptation path of SNR scalability as the predefined model when a specific spatial and temporal resolution is selected. Although the computation is less than a transcoder or decoder, the computation is still an issue when the amount of clients is enormous.

3. MULTIDIMENSIONAL ADAPTATION SELCTION

3.1. Adjustment of Video Adaptation in SVC

Video adaptor plays an important role for the applications in scalable video framework when the video parameters setting of the decoder are not the same as those of the encoder as shown in Fig. 1. In this case, video adaptor has to adjust the appropriate video parameter settings while selecting multidimensional video adaptation among the compressed scalable video stream. The decisions of video parameter settings are selected according to the equipments' constraints and the feedback information expresses the configurations of each client. In order to achieve the emerging requirements of better flexibility on scalable video stream, the higher flexibility of scalable video stream is required. For multidimensional video adaptation, the video adaptor needs more complicated selection rules to select the approximate video parameter setting. Fig. 2 shows the block diagram of video adaptor of SVC. The configurations are defined from the manual mode in original adaptor. The original video adaptor engine of SVC, which takes in two inputs: a scalable video stream and configurations as shown in Fig. 2(a). The original adaptor extracts the adapted video stream according the configurations as user requested in advance. The configurations provide the decisions of video parameter settings by a predefined model, and specify how to adapt a scalable video stream.

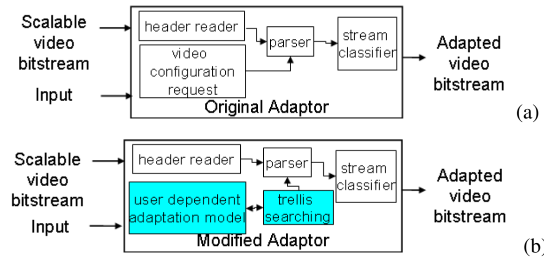

Fig. 2. (a) original adaptor engine, (b) modified adaptor engine.

Many video streaming applications 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 is not defined by each user. In this paper, we build a user dependent model to select the appropriate decision of multidimensional video adaptation selection. Moreover, the preference of each user is important for video adaptor to make a decision for multidimensional adaptation selection. 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. We use the feedbacks as the inputs to adjust the preference decision and statistical learning scheme applied in the modified video adaptor engine when selecting the video parameters. 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. After these adjustments, an appropriate video parameter setting which can fit the subjective preference of each user is selected for scalable video decoder. The proposed adaptor engine, based on the inputs, yields the adapted scalable video stream automatically.

3.2. Design Concept of Objectivity-Derived Algorithm

The basic design concept of our multidimensional adaptation selection scheme among the various video parameters is based on the concept of objective evaluation model. The objectivity-derived algorithm derived user dependent selection model for multidimensional video adaptation is described as our work [8]. For our observation, the relationship between resource and user score of the video parameter setting has the same trend as RD curve in single dimensional SNR scalability on specific spatial-temporal resolution. When the resource is increasing, the user score increases and then saturates according to the user point of view. For this reason, the user score can be described as a RD-like curve for a specific spatial-temporal resolution trade-off.

For different combinations of spatial and temporal scalabilities, different sensory feeling of each user is not easy to represent for evaluating the relationship of spatial and temporal scalabilities. To solve this problem, the design concept of our scheme is to extend the trend of objective evaluation tool, the RD-like curve, as the multidimensional selection model. We use single a RD-like curve to model the SNR scalability and take the independent RD-like curves to model the complex relationship of different spatial and temporal scalabilities. Then, we shift and scale the RD-like curves to adjust the switching bitrate points to match the video parameters of each user's preferences. The different RD-like curves cross the switching bitrate points. These points are useful to provide the prediction of the adaptation operation with different spatial and temporal resolution.

By doing this, we do not have to decode the compressed scalable video stream. The computation overhead is less than prior works and can be easily used when the amount of clients is enormous. We also do not have to derive an impartial model for the complex intuitions of each user and impartial evaluations about the visual impression of the relationship between different spatial-temporal resolutions. Each RD-like curve models the dimension of SNR scalability in multidimensional adaptation directly. For the complex relationship between spatial and temporal scalabilities, different RD-like curves model the SNR scalabilities of each different spatial-temporal resolution. Then, each emulated curve is shifted and scaled to the right place according to the input preferences of each user. Our proposed model can provide more appropriate adjustments of video parameters for each user.

3.3. Soft Decision Scheme

After objectivity-derived algorithm is derived, optimization fitting of proposed model for each user is the next important things. With the newly derived objectivity-derived algorithm, the optimization fitting of each user becomes: Use the input video parameter settings to make the And the residue matching error ε_m should be minimized.

$$\varepsilon_m = \left| M - \sum_{j=0}^{R_m-1} user\_score_{highest} \right|, \tag{1}$$

where M is the suppose target model, $user\_score_{highest}$ is the highest user score of the updated user dependent model which is adjusted according to the feedbacks.

One straight solution to the optimization problem is to adjust the parameters of objectivity-derived algorithm to approximate video parameter settings at every step.

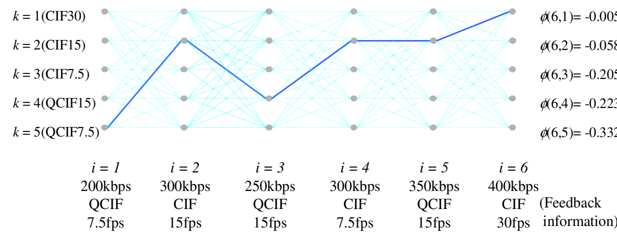

Fig.3. Trellis searching example.

However, it tries to approximate the input using a closest video parameter setting at each search step without looking ahead of future steps. It is easily to trap into the local solution.

In this paper, we propose an alternative soft decision optimization algorithm, called the trellis searching algorithm, to solve the optimization problem described in Eq. (1). The use of the trellis search can also be found in the applications of digital filter design [9]. We use an example to illustrate the proposed trellis searching algorithm. Suppose that the maximum iteration number is restricted to R_m = 6, and 5 possible candidate video parameter settings. First of all, let $C(Set)$ denote the number of the candidate video parameter settings, and each distinct video parameter setting in the set is expressed as $r(k)$, for $1 \leqq k \leqq C(Set)$.

In the trellis searching algorithm, there are $C(Set)$=5 states in each step. For kth state ($1 \leqq k \leqq C(Set)$) of ith search step, we use the cumulative candidate video parameter setting, $\Phi(i,k)$, to calculate the approximation model of each user in the kth state up to the ith step. That is,

$$\Phi(1,k) = r(k) \quad for\ all\ k, \tag{2}$$

A path in the trellis, which leaves the kth state at ith step and enters the kth state at $(i+1)$th step, corresponds to an operation of adding $\Phi(i,k)$ by $r(k')$. Then, the $\Phi(i+1,k') = \Phi(i,k) + r(k')$.

Conceptually, the whole process is similar to the trellis decoding of convolutional code: The trellis searching algorithm involves calculating and minimizing the difference between the suppose target model M and $\Phi(i,k)$ for all k at each search step i. To be specific, $\Phi(i+1,k)$ is determined in the way that

$$\Phi(i+1,k) = \min\{ |\Phi(i,k^*) + r(k) - M| : 1 \le k^* \le C(Set)\} \tag{3}$$

Then, the selected path is denoted as the surviving path. Note that we have to calculate all the cumulative candidate video parameter settings $\Phi(i,k)$ for all k before moving to the $(i+1)$th step until the maximum iteration number is reached ($i = R_m$).

After calculating the cumulative candidate video parameter settings for all states at the last search step, i.e., $\Phi(R_m,k)$ for $1 \leqq k \leqq C(Set)$, the next procedure for the trellis searching algorithm is to determine the global result.

Next, we can determine all the video parameter settings by tracing from the state, whose corresponding $\Phi(R_m,k)$ is best approximation of M, along its surviving path backward.

In this example, the procedure for trace back is illustrated in Fig. 3. All the surviving paths for each state at each step are represented by the dash line. $\Phi(6,1)$ is selected as the global result. We can thus determine the global surviving path of the trellis searching algorithm, as marked by the solid line in Fig. 3. In this case, $\Phi(6,1) = r(5) + r(2) + r(4) + r(2) + r(2) + r(1)$, which is the best approximation of M generated by the proposed trellis searching algorithm. After the trellis searching algorithm, the model is adjusted until it is converged. This scheme can find the most approximate model of each user is derived more quickly and precisely.

Compared with [7], our proposed model is more flexible to feet the user preference where the model of weighted PSNR work is limited according to PSNR curve. Second, we don't have to compute any extra information in advance for generating the model of adaptor. Sometimes the PSNR curves of all sequences are not available in advance in real applications. Finally, our proposed trellis optimization scheme also can overcome eliminate the uncertainty of the user as described in next subsection, which not discussed in presented literatures.

4. SIMULATION RESULTS

In this section, we implement our proposed modeling methods in video adaptor of JSVM [10]. The spatial resolution includes QCIF and CIF size. The temporal frame rate of QCIF size are 7.5 frames per second (fps) and 15fps, and the CIF size are 7.5fps, 15fps and 30fps, respectively. We adopt the single stimulus continuous quality evaluation (SSCQE) recommended by ITU-R standard with some revision to simulate adaptation of spatial and temporal scalability. 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.

The prediction accuracy of a predefined model means that we choose the most popular selection of all subjects at the selected bitrate. However, each user has their preference about the video parameters. A generic model is not appropriate to fit the variable preferences of every user. From the simulations, the proposed method provides a user dependent adaptation model for video adaptation application. Compared with a generic model, the R-D like curves can be adjusted as different models to reflect the user's subjective preference when different test sequences and users. The proposed method improves the accuracy prediction rate from 75% to 94% from minimum available bitrate to 1024kbps of test sequences by counting the number of subjects. In the experimental results, we see that the proposed model is intuitive and reflect the user's preferences adequately.

5. CONCLUSIONS

In this paper, we propose a scheme to match subjective quality evaluation of multidimensional selection for serving scalable video applications. This scheme can characterize the relationship between spatial, temporal and SNR scalability of each user's intuition. We use an objectivity-derived algorithm to derive the user dependent selection model for making the multidimensional scalability decision. Because we analyze the user preference according to the compress-domain data, this scheme can be used in video proxy or gateway without much computation overhead. Besides, this work considers the user preference, content information and objective consideration at the same time. The proposed method can achieve 94% accuracy prediction rate of when the switching bitrate point is stable. The proposed video adaptor can utilize the bandwidth in a more efficient way and a better subjective visual quality can be achieved. This makes the proposed scheme an easy and direct solution in practical use of video adaptor for serving scalable video applications.

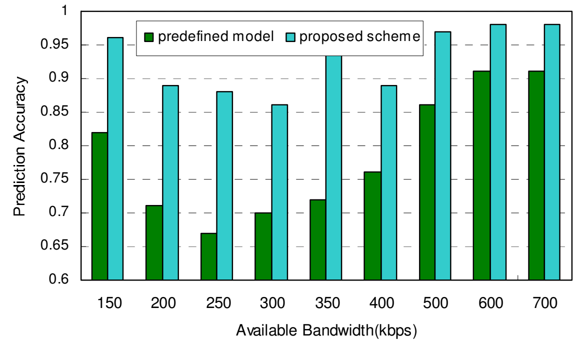

Fig. 4. Prediction Accuracy comparison.

6. REFERENCES

[1] S.-F. Chang and A. Vetro "Video adaptation: concepts, technologies, and open issues," *Proc. IEEE*, 2005.

[2] J. Xin, C.-W. Lin, and M.-T. Sun, "Digital video transcoding", *Proc. IEEE*, vol. 93, no. 1, pp. 84-97, Jan. 2005.

[3] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 11, pp. 332–344, 2001.

[4] "Scalable Video Coding Working Draft," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-P201, 2006.

[5] T. C. Thang, Y. J. Jung, and Y. M. Ro, "Effective adaptation of multimedia documents with modality conversion", *EURASIP Sig. Proc. Image Com. Journal*, Vol. 20, Issue 5, pp.413-434, 2005

[6] Y. Wang, M. v. d. Schaar, S. -F. Chang, and A. C. Loui, "Classification-based multidimensional adaptation prediction for scalable video coding using subjective quality evaluation," *IEEE Trans. Cir. Syst. Video Technol.*, pp. 1270-1279, Oct. 2005.

[7] T. C. Thang, D. S. Lee, J. W. Kang, and Y. M. Ro, "Optimal multi-layer adaptation of SVC video over heterogeneous environments," IEEE Int'l Conf. Wireless Commu. Mobil Compu., pp. 308-313, Aug. 2007.

[8] C.-H. Pan, I-H. Lee, S.-C. Huang, C.-J. Lian, and L.-G. Chen, "A Quality-of-Experiment Video Adaptor for Scalable Video Application", *IEEE Trans. Cons. Elec.*, pp. 1130-1137, Aug. 2007.

[9] A.-Y. Wu, I-H. Lee, and C.-S. Wu, "Angle quantization approach for lattice IIR filter implementation and its trellis de-allocation algorithm", *IEEE Int'l Conf. Accous. Speech Signal Proc.*, vol. 2, pp. 673-676, Apr. 2003.

[10]"JSVM 2.0 Software," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-O203, 2005.