Robust Video Transmission with an SNR Scalable H.264 Codec

M. Mahdi Ghandi and Mohammed Ghanbari

ESE Department, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK
{Mahdi, Ghan}@essex.ac.uk

This paper proposes a SNR scalability method based on the H.264/AVC video coding standard. The method partitions the compressed video data into layers with different quantization precisions. The base layer contains more important parts of the video data while the enhancement layer delivers the additional data to improve the video SNR quality. Therefore, it can be considered as an alternative method to the data partitioning technique that is already supported by the standard. By applying better error protection methods to the base layer, the delivered video is guaranteed to meet a minimum quality when transmitted over error prone environments. Simulation results show that the bitstreams generated by our layering method have better quality when transmitted in high erroneous channels compared to data partitioning.

1 Introduction

The capacity of a communication channel is determined by its bandwidth and its signal-to-noise ratio. For a digital user, these parameters determine the bit rate and the probability of error and so affect the achievable quality of service.

In recent multimedia systems especially wireless communication applications, bandwidth is still a limiting factor. Hence, video compression techniques are a crucial part of these applications. The H.264/AVC [1] video coding standard, proposed by the Joint Video Team (JVT) of ITU-T and ISO/IEC experts, achieves a significant improvement in the compression efficiency relative to other existing standards [2], [3], [4]. This makes the H.264 a serious contender for all multimedia applications.

However, due to the error-prone nature of wireless communication channels, the probability of error could be higher than in other channels. Therefore, receiving the correct data is not guaranteed. This implies a need for useful operation of robust video coding algorithms. The error resilience techniques supported in the H.264 such as slice structure, data partitioning and Flexible Macroblock Ordering (FMO) have been designed to address this issue and make the generated bitstreams more robust to transmission errors [5]. The data partitioning (DP) is an effective layering technique which partitions the compressed data in separate units of different importance. To improve the transmission efficiency, higher error protections are applied to the more important units of the coded data. In the recent video communication systems, layered coding with transport prioritization is the most popular and effective error resilience scheme [6].

Scalability is also considered as a powerful layering method to adapt the unequal error protection techniques [7]. It is an important feature of the recent video coders [8], [9] and has many applications in video streaming over wired and wireless communication channels [9]. Although the first version of H.264 [1] does not support scalability, it is listed on the work plan as an important tool that should be supported by the standard. Scalability is to partition a video bitstream into layers such that the base layer is an independent bitstream. The base layer is decodable into a video sequence with reduced quality (SNR) or spatial/temporal resolution. Enhancement layers provide additional data necessary for video reproduction with higher SNR or spatial/temporal resolution. SNR scalability, first proposed by Ghanbari [15], is to quantize the DCT coefficients to different levels of accuracy by using different quantization steps (determined by QPB and QPE). Therefore, the resulting streams have different quality (SNR) levels.

There are a number of proposals for scalability to be added to H.264 in the literature [11], [12] and the submitted contributions [13], [14]. In [11] and [12] the proposed scalability schemes apply only one Motion Estimation (ME) stage and so are not flexible to efficiently use layers data in prediction. Therefore, despite the benefit of low complexity, they suffer from lack of coding efficiency. In [13] a spatial scalability scheme is proposed that has a separate ME in the enhancement layer, but does not efficiently use the available motion data of the base layer. In [14] the efficiency of the proposed wavelet based SNR scalability method strongly depends on the contents of the sequence.

We have proposed a SNR scalability method based on H.264/AVC, which consists of an independent ME process to generate the enhancement data. To take advantage of the motion information of both layers, three different modes are designed for the enhancement Macroblock (MB) prediction. These modes make the coder flexible to maintain the efficiency in different coding situations. The MV coding method and the context models for the enhancement layer have small modifications, and the bitstream structure is adapted to achieve better compression efficiency.

On the other hand, we have simulated the transmission of the scalable bitstream through a noisy channel and compared it to the non-scalable bitstream with data partitioning. We have applied unequal error control techniques to protect the base layer more than the enhancement layer. Simulation results show that in the presence of unequal error protection, the scalable bitstream has a significant improvement in quality for higher bit error rates. However, due to the bit rate overhead of the scalable bitstream, the resulting quality in the lower bit error rates are somewhat worse than the non-scalable one.

The remainder of this paper is organized as follows. Section 2 gives a brief description of the proposed scalable method. In Section 3 the evaluation methods and simulation results are described and finally, Section 4 concludes this paper.

2. SNR Scalable H.264 Codec

The block diagram of the proposed scalable coder is illustrated in Fig. 1. The base layer is coded exactly the same as the standard non-scalable coder. The data of the current picture in the base layer is available for coding the enhancement layer as well as the previously coded pictures in the base and the enhancement layers. Therefore, to take advantage of all available information to improve coding efficiency, there are three prediction modes for the enhancement layer.

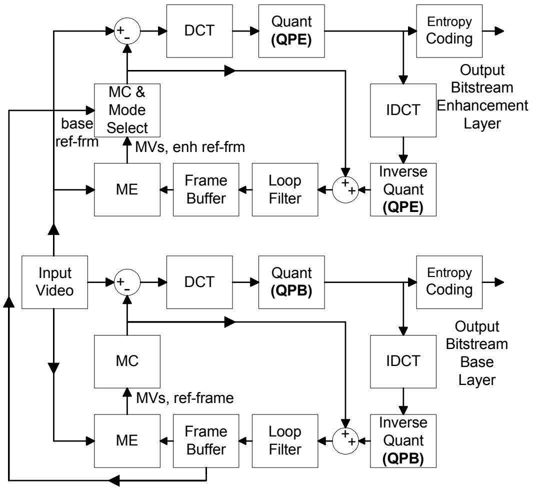

Fig. 1. The block diagram of the proposed scalable coder

2.1 Enhancement Prediction Modes

For the coding of every inter-coded block in the enhancement layer, it is firstly predicted by one of the upward, direct or forward prediction modes. In the upward mode no MV is sent for the block and the prediction is made by zero MVs. The reference picture is one of the base layer reconstructed pictures. This mode is useful especially when the base layer MB contains enough information and a small residual data is sufficient to represent the enhancement data. Furthermore, this mode stops a probable drift when the enhancement layer references have been received incorrectly due to an error. Additionally, in some situations there is a need to insert intra pictures in the base layer in order to have a random access point. In these cases encoder could be limited to select only upward mode (as well as intra modes) to support the random access in the enhancement layer as well.

In the direct mode similar to the upward mode no MV is sent, but the prediction is made by the MVs equal to those of the corresponding base layer pictures. The direct mode reference picture is one of the previously coded pictures of the enhancement layer. Since the moving objects in the base and the enhancement layer are the same, there is a reasonable correlation between the base and the enhancement MVs. Our simulations show that from 20 to 50 per cent of these MVs (depending on the picture contents and the quality difference between the base and the enhancement layer,) have almost same values and in those cases, it is more efficient not to send further MVs for the enhancement blocks, and hence direct mode would be selected.

In the forward mode, the new set of MVs generated by an independent ME process, is sent and the reference picture is among the previously coded pictures in the enhancement layer. This mode is selected more often when the quality difference between the base and the enhancement layer is larger and hence the base layer does not contain enough information (including motion and residual data). However, since there are still some correlations between the base and the enhancement layer, we modified the enhancement entropy coding methods in order to have more efficient compression.

As well the above inter modes, an enhancement MB can be coded in intra mode which is selected very rarely since it is not efficient. Furthermore, a MB can be skipped and in this case no information is sent for that MB. Some other scalable methods [8], [13] have another mode called bidirectional in which the predicted block comprises of an average of forward and upward predictions. However, in the simulation results we will show that including this mode which adds more complexity to the encoder, will not improve the coding efficiency.

In our coder, a Lagrangian optimization process selects the proper mode for every block. Using the selected modes, motion compensation, block transformation and quantization are performed. The resulting MB headers, MVs and the residual data should be coded and sent. The coding methods and the bitstream structure have small modifications described in the following section.

2.2 The Enhancement Layer Structure and Coding

In H.264 every inter-coded, 16x16 pixel MB can be partitioned into various block sizes and shapes illustrated in Fig. 2. The partitioning choice of a MB into 16x16, 8x16, 16x8 or 8x8 blocks is determined by *mb-type*. In 8x8 mode (i.e. *mb-type* 3) each of the blocks can be further divided independently into 8x8, 8x4, 4x8 or 4x4 sub-partitions determined by *sub-mb-type*. Every macroblock partition (but not sub-partition) shown on the top of Fig. 2, could have a different reference picture determined by *ref-idx* and each small block contains a separate MV.

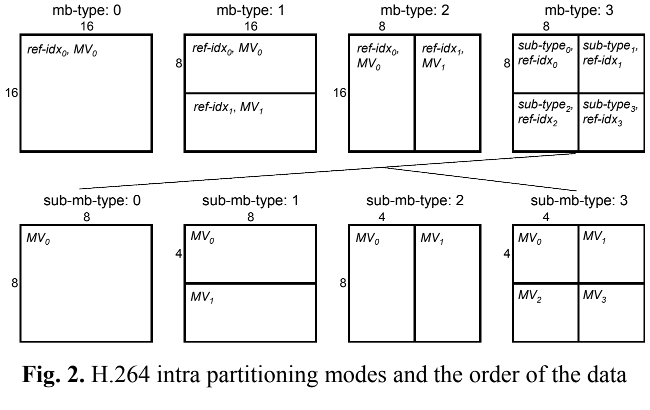

Fig. 2. H.264 intra partitioning modes and the order of the data

In our implementation for the enhancement layer, in order to address the prediction modes we have changed the semantic of the *ref-idx* such that 0, 1 and 2 values denote forward, upward and direct modes respectively for the reference picture 0, and in the same way the higher values point to other reference pictures. In direct and upward modes since there are no MVs and they are completely specified in the base layer, there is no need to send *sub-mb-type*. Therefore, we moved the appearance order of *ref-idx* before *sub-mb-type* in the syntax structure to prevent sending unnecessary data. One can modify the bitstream syntax to another structure. For example add more possible modes in *mb-type* and *sub-mb-type* instead of changing the semantic of *ref-idx* and it may be more sensible. However, it does not have a significant effect on the coding efficiency which is the aim of our simulation.

In the base layer of H.264 coder, for coding an MV, it is firstly predicted from the MVs of neighbor blocks (generating PMV) and then the difference between the original MV and PMV (i.e. MVD) is calculated and coded. In the enhancement layer, when a neighbor block is in upward or direct mode or intra coded, it does not have MV. In these cases the MVs of the corresponding base layer blocks are used for prediction. This will enhance the MV prediction accuracy and hence improves the coding efficiency of MVD.

To code the MVD in CABAC mode, the sign and absolute value of the MVD are coded separately. For arithmetic coding the sign of MVD in the base layer, since it is statistically almost equal to be negative or positive, equal probability model is used. However, in the enhancement layer, MVs have a correlation to the corresponding base layer MVs and so there are 4 different probability models for sign coding (two for horizontal and two for vertical MVD) addressed by an index. The context index of MVD sign is determined by:

$$\begin{aligned} & mvd\_sign\_context\_index = \\ & 2 \times hv + sign(base\_mv(hv) - PMV(hv)) \\ & ; sign(x) = (x < 0) ? 1 : 0 \end{aligned} \tag{1}$$

where *hv* is 0 for horizontal and 1 for vertical and *base_mv* is the corresponding base layer MV. It can be observed that this context depends on the base layer MV which has a high probability to have a value near the enhancement layer MV.

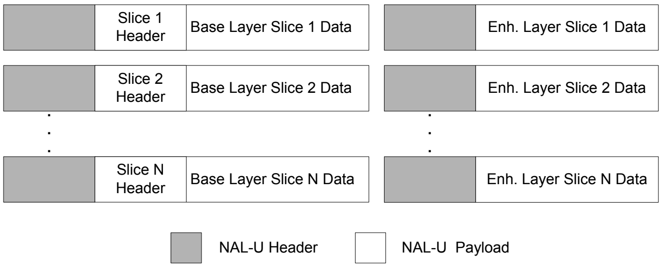

Fig. 3. The scalable bitstream NAL Unit distribution

In H.264, the coded video data (coded slices) are placed in Network Adaptation Layer (NAL) units to facilitate the delivery of the data to the underlying transport layers. In the scalable bitstream every enhancement layer slice is placed after its corresponding base layer slice in a separate NAL unit as shown in Fig. 3. By this method of NAL unit distribution, duplicate sending of slice headers is prevented as well as providing the ability to control the Unequal Error Protection (UEP). Note that when the Data Partitioning (DP) mode is enabled (in either scalable or non-scalable bitstreams) each slice is further divided into three different NAL units (classified according to importance,) providing the capability to manage the UEP.

3. Simulation Results

The encoder and decoder of the proposed method have been implemented using the standard JVT codec software version 7.3. Several tests have been conducted to verify the Rate Distortion (R-D) performance of the proposed scalable coder in the noise-free situations. In Fig. 4 the R-D curves of the Foreman test sequence are illustrated. The scalable coder has been tested for three different values of dQP (QPB-QPE). It can be seen that when more bit rate budget is allocated to the enhancement layer (larger dQP) the efficiency of the scalable coder is better and closer to the non-scalable one. The reason is that in SNR scalability, what is actually coded in the enhancement layer is the quantization distortion of the base layer [10]. Therefore when dQP is small, the distortion coefficients are normally smaller than the enhancement quantization step size and would not be coded (re-quantized to zero). Hence, the enhancement layer does not improve the picture quality noticeably while sending a reasonable amount of addressing data. On the other hand at high dQPs, the quantizer step size of the base layer is large and hence the second layer efficiently codes any residual base layer quantization distortion. It should be mentioned that in these cases the quality of the base layer is poor.

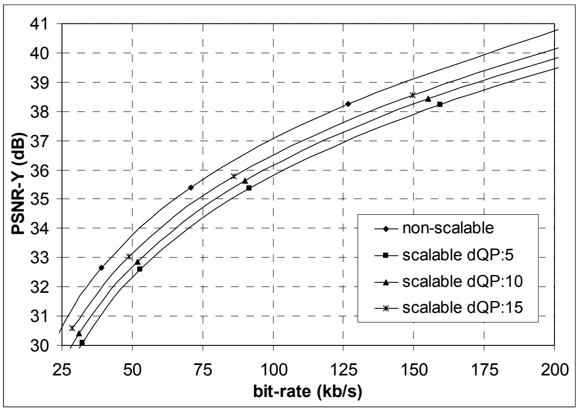

Fig. 4. Error-free R-D performance of the scalable coder (with fixed QPs) in three different conditions compared to the non-scalable one. Foreman QCIF@10Hz, 133 frames coded

To verify the efficiency of the proposed set of enhancement prediction modes, different combinations have been examined. In this experiment, in order to show that the bidirectional mode does not have an improvement and hence should not be included in the available modes, a separate Lagrangian optimized ME was performed for this mode to obtain the best possible gain. Note that this would significantly increase the encoder complexity. Table 1 shows the bit rate overhead of the scalable scheme compared to the non-scalable one in various combinations of modes in two scenarios, for the "Foreman" sequence. From the table it can be seen that adding the bidirectional mode has increased the overhead (as well as the complexity) of the scalable coder. However, adding the direct mode has reduced this overhead (by reducing the overhead of enhancement MVs,) and hence has been selected in our proposal.

Table 1. Overhead of the scalable coder in different scenarios of available enhancement prediction modes, Foreman QCIF@10Hz

Available modes	% overhead QPE:28, QPB:32	% overhead QPE:28, QPB:38
Forward and Upward (FU)	39.8	23.6
FU and Bidirectional (FUB)	39.7	25.2
FU and Direct (FUD)	**34.2**	**21.9**
FUBD	36.0	23.1

A series of tests is performed to evaluate the scalable bitstream robustness against the channel noise and compare it to data partitioning. To simulate the channel errors, an Elliot-Gilbert two level error model [10] has been used to introduce bit errors on the bitstreams. In the decoder, before any assessment, the corrupted parts of the video data are concealed using the method of [16]. To make the comparisons_fair, the test sequence is first encoded in a non-scalable scenario with enabled data partitioning. The scalable coder is then forced to adjust the base layer bit rate equal to the first partition (DPA) of the non-scalable bitstream. This is done using our Lagrangian optimized rate-control technique [17]. The total bit rates of both scalable and non-scalable bitstreams are also controlled to be the same.

Fig. 5 shows the test results for the foreman sequence. In this test, every row of MBs has been adjusted as one slice (i.e. 9 slice per frame). The average PSNRs in different bit error rates are shown for non-scalable bitstream when the DPA part of the bitstream is protected and considered as error free. For the scalable bitstream, the protected part is the base layer. Data partitioning feature for the scalable bitstream was disabled. It is obvious that in low bit error rates (error free channel) the scalable bitstream has lower quality as a result of the overhead. However, in higher error rates, the scalable bitstream has much better quality. The reason is that in scalable bitstream, the base layer is independently decodable. However, in the non-scalable bitstream, the loss of data even in the non-important partitions causes a propagating drift into the sequence.

In Fig. 6 the results of the same test is depicted when the unequal error protection technique is applied to both bitstreams using hierarchical QAM 18], [19]. It was simulated in a 64 QAM scenario with α (the distance separating two points in the constellation diagram) equal to 4. Note that this method is just an example among several techniques of error protection. The problem of how to efficiently apply error protection is another issue and is the subject of future research. From the figure it can be observed that the scalable bitstream has a significant improvement in the lower values of channel SNR compared to the non-scalable bitstream. It should be added that in this particular test, the DPA bit rate (in the most efficient mode of the coder) was 33% of the bit rate whereas in general case, adjusting the DPA bit rate would significantly decrease the coder efficiency. However, the scalable coder is always able to control the base layer bit rate flexibly with limited efficiency degradation (see Fig. 4).

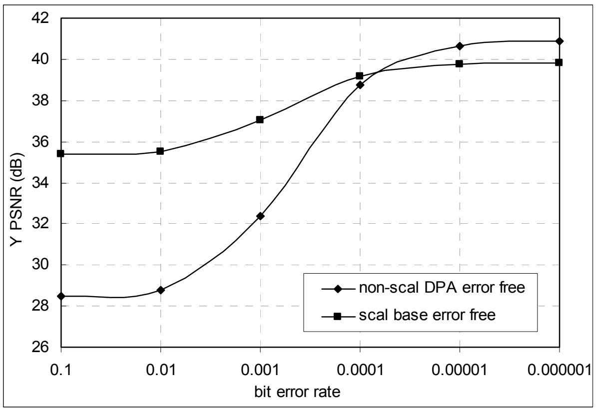

Fig. 5. Average output quality of the receiving scalable and non-scalable (data-partitioned) bitstreams, with error protection for DPA and the base layer. Foreman QCIF@10Hz, 200Kbits/Sec, DPA and base layer bit rates are 33% of total bit rates

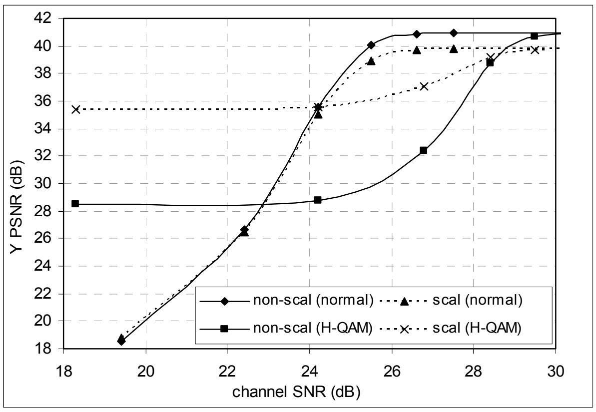

Fig. 6. PSNR vs. channel SNR, Foreman QCIF@10Hz, 200Kbits/Sec, scalable and non-scalable with equal error protection (normal) and unequal error protection using hierarchical QAM α: 4 (H-QAM)

4. Conclusion

We have proposed a new SNR scalable coder based on the H.264/AVC video coding standard. A new set of prediction modes is proposed for the enhancement layer as well as modifications to the coding and bitstream structure. The scalable coder has an acceptable R-D performance especially when more bit budget is allocated to the base layer. Furthermore, compared with the non-scalable ones, scalable bitstreams in conjunction with unequal error protection are more robust to the channel errors.

Acknowledgement

This project is supported by Engineering and Physical Sciences Research Council (EPSRC) of the UK. The authors would also like to thank Dr. Ernesto Luiz Andrade Neto for his valuable contributions.

References

1. Joint Video Team (JVT) of ISO/IEC MPEG, ITU-T VCEG: Final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC). JVT-G050, 2003
2. Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard and Ajay Luthra: Overview of the H.264/AVC video coding standard. IEEE trans. on circuits and systems for video technology, Vol. 13, No. 7, 560-576, July 2003
3. Daniel Alfonso, Daniele Bagni, Danilo Pau and Antonio Chimienti: A Performance analysis of H.264 video coding standard. Picture Coding Symposium, Saint-Malo, France, April 2003
4. Anthony Joch, Faouzi Kossentini and Panos Nasiopoulos: A Performance analysis of the ITU-T draft H.26L video coding standard. Int. Packet Video Workshop, Pittsburgh, PA, April 2002
5. Thomas Stockhammer, Miska M. Hannuksela, Thomas Wiegand: H.264/AVC in Wireless Environment. IEEE Trans. on Circuits and Systems for Video Technology, Vol. 13, No. 7, July 2003
6. Y. Wang, Q. F. Zhu: Error Control and Concealment for Video Communication: A Review. Proc. IEEE, Vol. 86, 974-997, May 1998
7. Michael Gallant, Faouzi Kossentini: Rate-Distortion Optimized Layered Coding with Unequal Error Protection for Robust Internet Video. IEEE Trans. on Circuits and Systems for Video Technology, Vol. 11, No. 3, March 2001
8. International Telecommunication Union (ITU): Recommendation H.263: video coding for low bit rate communications. Feb 1998
9. ISO/IEC MPEG: Coding of audio-visual objects – Part 2: Visual. ISO/IEC 14 496-2 (MPEG-4 Visual Version 1), Apr. 1999
10. Mohammed Ghanbari: Standard Codecs: image compression to advanced video coding. IEE Telecommunications Series, 2003
11. Yuwen He, Feng Wu, Shipeng Li, Yuzhuo Zhong and Shinqiang Yang: H.26L-based fine granularity scalable video coding. IEEE int. symposium on circuits and systems, Vol. 4, 548-551, Phoenix, USA, May 2002
12. Till Halabach, Thomas R. Fischer: SNR Scalability by transform coefficient refinement for block-based video coding. Visual Communications and Image Processing Conference, Lugano, Switzerland, July 2003
13. Lukasz Blaszak, Marek Domanski and Slawomir Mackowiak: Spatio-Temporal scalability in AVC codecs. ISO/IEC JTC1/SC29/WG11, Pattaya, March 2003

14. Heiko Schwarz, Detlev Marpe, and Thomas Wiegand: SNR-scalable extension of H.264/AVC. ISO/IEC JTC1/SC29/WG11, JVT-I032d1, San Diego, September 2003
15. Mohammed Ghanbari: Two-Layer coding of video signals for VBR networks. IEEE journal of selected areas in communication, Vol. 40, No. 9, 771-781, June 1989
16. Ye-Kui Wang, M.M. Hannuksela, V. Varsa, A. Hourunranta and M. Gabbouj: The Error Concealment Feature in the H.26L Test Model. IEEE ICIP 2002
17. M. Mahdi Ghandi and Mohammed Ghanbari: A Lagrangian optimized rate control algorithm for the H.264/AVC encoder, ICIP 2004, submitted for publication.
18. M. Morimoto, M. Okada and S. Komaki: Robust mobile image transmission using hierarchical QAM. Elsevier image communication journal, No. 12, 127-134, 1998
19. B Barmada and E V Jones: Adaptive Mapping and Priority Assignment for OFDM, IEE conference of 3G mobile communication technologies, London, 2002