HYBRID RESOLUTION SWITCHING METHOD FOR LOW BIT RATE VIDEO CODING

Sang Heon Lee, Sang Hwa Lee, and Nam Ik Cho

School of Electrical Eng. and Computer Sciences, INMC, Seoul National Univ., Korea

ABSTRACT

This paper proposes a video coding method using hybrid resolution switching for low bit rate environments. The proposed method encodes I pictures in high resolution and B and P pictures in low resolution. The decimated inter-frame pictures are encoded in the usual H.264/AVC framework, and they are reconstructed to high resolution ones using motion information, interpolation filter, and some residual signals in high resolution. The proposed video coding scheme shows better compression performances in low bit rate environments than the traditional algorithms based on H.264/AVC. The resolution switching method increases the coding efficiency in low bit rates, mainly because the side information of motion estimation is reduced. It is expected that the proposed method can also improve higher bit rate cases, if some parameters optimization and deblocking filter are appropriately applied.

Index Terms— video coding, H.264, resolution switching, low bit rate.

1. INTRODUCTION

Many video applications are based on the real-time bit streaming services. Video contents are provided via networks and reconstructed as soon as the decoders receive them. However, except for digital TV systems, we have not enough transmission capacity to send high resolution video data. This situation increases the need of low bit rate video coding.

Generally, the side information is more important in low bit rate video coding since most of transformed residual signals become zeros by the large quantization steps. In other words, the amount of information for the transform coefficients is very small compared with the side information such as motion information, when the QP is set to high values for low bit rate coding. This can be certified by simple simulations with IBBP GOP structure and several fixed QP's. Table 1, 2, and 3 show the ratio of data quantity between residual signals and side information. The crew sequence in Fig. 3 was used in IBBP GOP structure of 100 frames. The tables show that side information have a greater portion of the total data as QP's increase. Therefore, reducing side information in low-bit rate coding can further increase coding efficiency. Also, the side information is related mainly with motion estimation processes. H.264/AVC video coding standard uses more complex motion estimation and motion compensation (MEMC) schemes to yield better coding efficiency [1, 2]. Thus, H.264/AVC systems have much side information to be transmitted for MC in the decoders.

Table 1. Generated bit ratio in H.264/AVC for I frame

QP	header bits/frame	texture bits/frame	$\frac{\text{texture bits}}{\text{header bits}}$
24	5963.9	273599.1	45.88
30	7933.3	121051.4	15.26
36	8044.9	53088.3	6.60
42	7195.6	22493.0	3.13

Table 2. Generated bit ratio in H.264/AVC for P frame

QP	header bits/frame	texture bits/frame	$\frac{\text{texture bits}}{\text{header bits}}$
24	13047.7	163570.1	12.54
30	9527.1	29776.7	3.13
36	5953.3	6884.2	1.16
42	3223.0	1849.1	0.57

Table 3. Generated bit ratio in H.264/AVC for B frame

QP	header bits/frame	texture bits/frame	$\frac{\text{texture bits}}{\text{header bits}}$
24	12754.4	111151.8	8.71
30	6612.6	10050.7	1.52
36	2511.1	861.7	0.34
42	792.1	111.9	0.14

This paper proposes a resolution switching method to reduce the side information in video coding. The motion estimation is performed in low resolution inter-frames so that side information on the MEMC is decreased. Of course, the MEMC process in low resolution may increase residual signals power. However, the high QP's in low bit rate environments quantize most of error signals to zero, and the residual signals and distortions do not increase so much. And Zeng and Venestanopulos[3] showed that higher coding gains can be achieved at low-bit rate JPEG image coding by downsampling original image and interpolating it after decoding. Thus, residual signal power in the proposed scheme is not highly increased in low bit rate video coding. Furthermore,

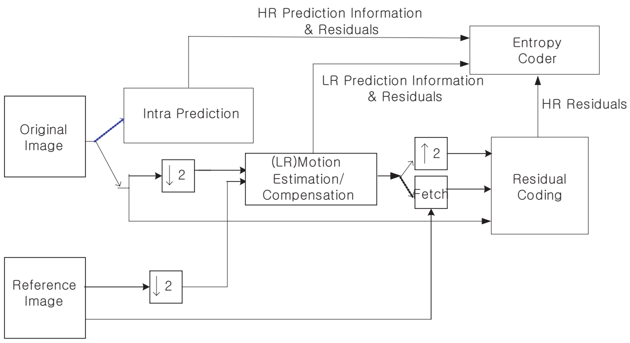


Fig. 1. Encoder block diagram.

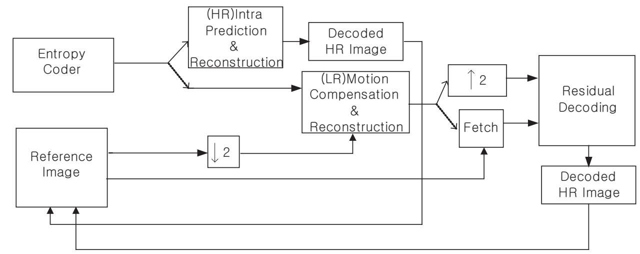


Fig. 2. Decoder block diagram.

to increase estimation accuracy from low resolution image to high resolution image, inter-layer prediction methods in SVC[4] spatial scalability are used. To compensate for loss caused by MEMC processes in low resolution, this paper also proposes some reconstruction methods to recover the original high resolution frames.

The rest of paper is organized as follows. The overall structure of resolution switching method is presented in Section 2. The prediction schemes of HR frames is described in Section 3, and the reconstruction of inter-frames in high resolution is explained in Section 4. Experimental results and conclusions are shown in Section 5 and 6, respectively.

2. RESOLUTION SWITCHING METHOD

Fig. 1 shows the block diagram of proposed encoder. Intra-frames (I pictures) are encoded in the original high resolution (HR) by H.264/AVC encoding algorithm. The HR information of intra-frames are well encoded and transmitted to the decoder for the reconstruction of HR frames. On the contrary, inter-frames (B and P pictures) are decimated and encoded by the usual H.264/AVC in low resolution (LR). Motion estimation and compensation are performed with H.264/AVC algorithms in low resolution. Residual signals between decimated input frame and motion compensated one are transformed, quantized, and entropy coded with H.264/AVC algorithms. After this encoding process in low resolution, we get the decoded LR frame of the input in the encoder. Using the decoded LR frame and motion vectors, the proposed algorithm predicts an HR frame by interpolation or motion fetch method for every block. Then, error signals between the original HR input frame and predicted HR one are transformed, quantized, and entropy coded. Note that the error signals are generated in high resolution. Finally, an R-D optimization scheme is introduced to determine whether the HR error signals are transmitted or not for every 32x32 high resolution block.

Fig. 2 shows the decoder block diagram of proposed video coding method. Intra-frames are decoded in high resolution by the same process of H.264/AVC. Inter-frames have two steps of reconstruction. First, the LR inter-frames are decoded using decimated reference frames, motion vectors, and LR residual signals in the H.264/AVC framework. Second, the HR inter-frames are predicted with interpolation and motion vectors as is the same process of encoder. The HR error signals are decoded and added to the predicted HR frames, which reconstructs the final HR inter-frames.

3. HR INTER-FRAME PREDICTION

The proposed algorithm predicts HR inter-frames from decoded LR ones. The prediction is performed using motion information estimated in low resolution and image interpolation. Each 16×16 macroblock in the decoded LR frames are interpolated to a 32×32 block or enlarged by fetching the corresponding 32×32 block in the HR reference frames. Each prediction method is described as follows.

3.1. Interpolation

The interpolation method uses 6-tap FIR filter, which is used in the H.264/AVC to estimate motion vectors of quarter pixel accuracy. It has the coefficients of $\{1, -5, 20, 20, -5, 1\}$. This filter is applied to each 16×16 macroblock in the decoded LR inter-frames, and makes 32×32 block in the HR frames. Block boundaries are padded by mirroring boundary values. This interpolation method can predict HR blocks well from LR blocks when the LR blocks are homogeneous. Furthermore, the decoder does not need additional side information about the interpolation filter. However, the interpolation method makes high prediction errors if the blocks have complex textures. The error signals between interpolated HR block and original HR one are encoded and transmitted to the decoder.

3.2. Motion fetching

Motion fetching method uses a motion vector and the corresponding reference indices which have been found in motion estimation process in low resolution. In encoding inter-frames in low resolution, motion vector and corresponding reference

Table 4. Bit strings for HR reconstruction modes.

Mode	Prediction mode	Sending residuals	Bit string
Mode1	Interpolation	sending residual	00
Mode2	Interpolation	No residual	01
Mode3	Motion fetching	sending residual	10
Mode4	Motion fetching	No residual	11

indices are stored for every 4×4 block. They are used to obtain motion vector and reference indices for 8×8 block in the predicted HR frame. The reference indices for 4×4 block in the LR frame are used as those of 8×8 block in the predicted HR frame. Motion vectors in the LR frame is doubled and assigned to those in the predicted HR frames because frames are doubled,

$$MV_x^H = 2MV_x^L, \text{ and } MV_y^H = 2MV_y^L, \qquad (1)$$

where MV_x^H and MV_x^L mean the x-axis motion vectors in high and low resolution, respectively. By using the transferred reference indices and doubled motion vectors, the 32×32 blocks in the HR frames are predicted. And the error signals between motion fetched HR block and original HR one are also encoded and transmitted to the decoder.

4. HR RECONSTRUCTION

The proposed algorithm predicts HR frames from the decoded LR ones. The error signals between predicted HR frame and original HR one should be transmitted to recover HR frame perfectly. The error signals are generated by two processes, interpolation and motion fetching. However, transmitting all error signals doesn't guarantee good coding efficiency. If the distortion between original HR frame and predicted HR one is not significant, it is more efficient not to send the HR error signals. From the point of R-D optimization, we determine whether the HR error signals are transmitted or not. R-D optimization technique in the H.264/AVC [5, 6] is exploited as

$$J = D + \lambda R. \qquad (2)$$

The R-D optimization process determines the prediction method out of interpolation and motion fetching as well as transmission of error signals. Since there are 4 modes to recover HR frames from decoded LR ones, additional information to indicate the mode is required to decode the proposed algorithm. The proposed algorithm transmits the side information with 2 bits. Table 4 summarizes binary bit strings for the modes of HR reconstruction. Those additional bits for a block do not increase overall transmission data because motion information to be transmitted is reduced in low resolution. Each bit string is encoded by the CABAC method [7, 8, 9].

5. SIMULATION RESULTS

The proposed algorithm is tested on JM 9.5 with high profile and IBBP GOP structure [10]. We selected high resolution test sequences of 4:2:0 format, such as two 4CIF sequences and two 640×480 ones, in order to prove the high compression performances in low bit rates. The 4CIF sequences CREW and SOCCER with 298 frames are used at SVC standardization, and tested for three bit rates, 300kbps, 450kbps, and 600kbps. The 640×480 sequences Ballroom and Race1 with 250 frames are used at MVC standardization, and tested for three bit rates, 256kbps, 384kbps, and 512kbps. Fig. 3 shows the coding results of proposed algorithm at low bit rates. In the case of 4CIF sequences, the proposed algorithm shows better coding efficiency than H.264/AVC until 600kbps. At low bit rate such as 300kbps, the proposed algorithm has coding gain higher than 0.7 dB (PSNR). In the case of 640×480 sequences, the proposed algorithm also shows better coding efficiency. However, in the experiments with ballroom sequence, the bit rates to have better performance become decreased to 384kbps. This results from complex scene structure of ballroom sequence compared with other sequences. The complexity of the scenes increases the power of residual signals during interpolation and reconstruction of high resolution frames. For the experiments with simple sequences such as race1, the proposed algorithm shows much better coding gain at higher bit rates over 512kbps. Simulation results with various sequences show that the proposed algorithm outperforms the usual H.264/AVC based coding in low bit rate environments. Also, the coding gain of proposed method is equal or slightly less than the usual H.264 based algorithms at higher bit rates. Further works should be continued to increase the bit rates to show better coding gain than the traditional approaches for the complex scenes.

6. CONCLUSIONS

This paper has proposed a video coding structure for low bit rate environments using hybrid resolution switching. The proposed method encodes I pictures in high resolution and B and P pictures in low resolution, which has reduced the side information in video coding. The inter-frames decimated to low resolution pictures are encoded in the usual H.264/AVC framework, and reconstructed to high resolution ones using motion information, interpolation filter, and residual signals between interpolated frames and original ones in high resolution. Simulation results show that the proposed algorithm has better coding efficiency in low bit-rate environments than the traditional algorithms based on H.264/AVC. The proposed method is suitable for the systems that require low bit rates such as video streaming services, real-time transmission of high resolution videos, storage media of high compression efficiency, etc. Further research is required in the interpolation method, deblocking filter, R-D optimization to send high resolution error signals, and adaptive resolution switching based on scene complexity.

7. REFERENCES

[1] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC)," JVT of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, 2003.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC coding standard," *IEEE CSVT*, vol. 13, pp. 560-576, 2003.

[3] B. Zeng and A. N. Venetsanopoulos "A JPEG-based interpolative image coding scheme," *IEEE ICASSP*, vol. 5, pp. 393-396, 1993.

[4] "Joint Draft9 of SVC Amendment with proposed changes," ISO/IEC JTC1/SC29/WG11 & ITU-T SG16 Q.6 Doc. JVT-V202, Jan, 2007.

[5] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no.6, pp. 74-90, Nov. 1998.

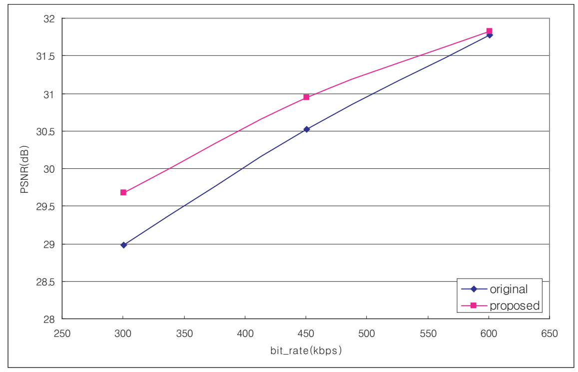
[6] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," *IEEE ICIP*, vol. 13, pp. 542-545, 2001.

[7] D. Marpe, G. Blattermann and T. Wiegand, "Improved CABAC," ITU-T SG16/Q.6 Doc. VCEG-018, 2001.

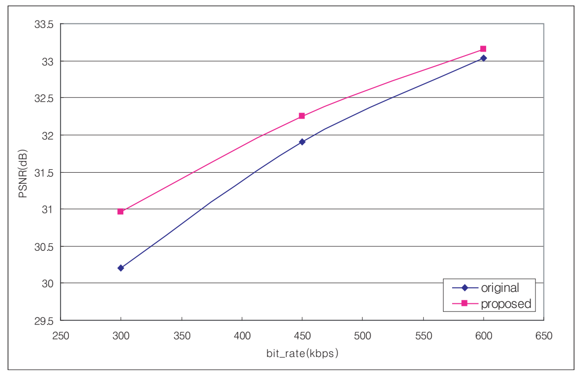
[8] D. Marpe, H. Schwarz, G. Blattermann and T. Wiegand, "Final CABAC cleanup," ISO/IEC JTC1/SC29/WG11 & ITU-T SG16 Q.6 Doc. JVT-F039, Dec. 2002.

[9] D. Marpe, H. Schwarz and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE CSVT*, vol. 13, pp. 620-636, 2003.

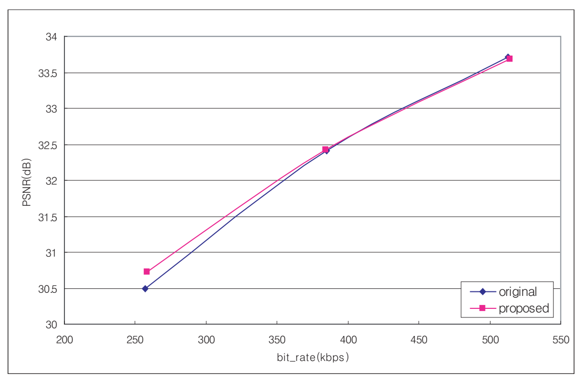
[10] http://iphome.hhi.de/suehring/tml/download/, H.264 JM reference model.



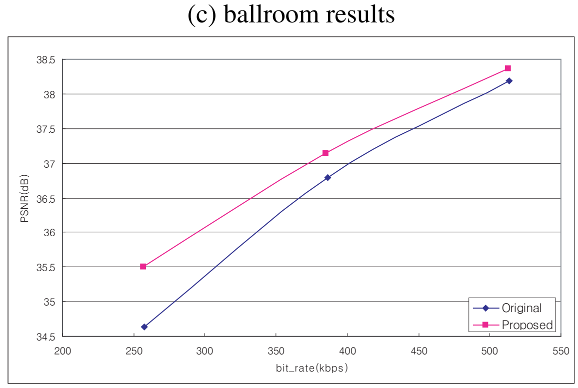
(a) soccer results



(b) crew results



(c) ballroom results



(d) race1 results

Fig. 3. Coding results in low bitrate environments.