An SNR Scalable Video Coding using Linearly Combined Motion Vectors

Chang-Hoon Ryu, Byoungjun Han, Kwang-Pyo Choi, Eung Sik Yoon, Keun-Young Lee
Department of Electrical and Computer Engineering, SungKyunKwan Univ.
300 Chunchun-dong, Jangan-gu, Suwon, Kyunggi-do, Republic of Korea
Tel: +82-31-290-7193, Fax: +82-31-290-7180
E-mail: faraday@freechal.com, formax@ece.skku.ac.kr

Abstract: There are increasing needs to deliver the multimedia streaming over heterogeneous networks. When considering network environments and equipment accessed by user, delivery of video streaming must be scalable. There are many kinds of scalable video coding: spatial, temporal, SNR, and hybrid. The SNR scalable coding produces video streaming, which has same temporal and spatial resolution, but different SNR quality with respect to layers. The 1-layer SNR scalable encoder produces SNR scalable video streams with ease. But, there is drift problem. Modified 1-layer approach does not have this problem but coding inefficiency, and is not MPEG-compliant. The present MPEG-compliant 2-layer encoder comes out to reduce coding rate. But it still use only base layer to encode whole layer. In this paper, we propose adaptive MPEG-compliant 2-layer encoder. Using linear combination algorithm, encoder use 1 motion vector to encode the sequences efficiently. By dong this, we can achieve the coding efficiency of SNR scalable coding.

1. Introduction

The scalable (layered) coding generates two or more bit-streams (layer): A Base Layer and one or more Enhancement Layers. Generally, The base layer has the header information and information of low resolution picture generally. The enhancement information has residual information to enhance low-level picture of the base layer.

The scalable coding is originated from robustness of multimedia transmission. [2] If network status is bad - the bandwidth of network is less than data rate of video sequence of full quality -, some of data packet will be disappeared in transmission. If the priority of transmission is allowed, low-level picture data will be at least transmitted. In transmission, the information of base-layer always must be received by decoder. Bit stream of the enhancement layer is not decodable on itself. The enhancement information enhances low resolution picture of the base layer.

There are many kinds of scalable video coding: spatial, temporal, SNR and hybrid. The spatial scalable coding generates bit stream of different spatial resolutions with respect to layers. The temporal scalable coding generates bit stream of different temporal resolution: ex) If decoder receive the bit stream of base layer and enhancement layer, 30 frame, or 15 frame(base layer) will be decoded. We can make the bit stream of different visual quality while we make spatial and temporal resolution same with respect to each layers. The object visual quality is measured as PSNR (Peak-to-Peak Signal to Noise Ratio). The SNR (Signal to Noise Ratio) scalable encoder generates different 'PSNR' quality sequence with respect to each layers.

Several kinds of SNR scalable coders have been presented so far; 1-layer, modified 1-layer, 2-layer coding structure, and so on. Main different feature among them is how to include the information of enhancement layer in reconstructing picture at frame memory of encoder and decoder.

$$PSNR = 10\log_{10}\left(\frac{MaxValue^2}{MSE}\right)$$

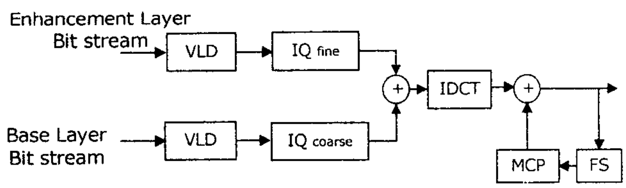
$$= 10\log_{10}\left(\frac{255^2}{\frac{1}{N}\sum\sum\left(Y_{ref}(x,y) - Y_{org}(x,y)\right)^2}\right)$$

In this paper, we propose a SNR scalable encoder, which has reduced rate by using linearly combined motion vector (LCMV). This paper is organized as follows. First, we give a brief description of the SNR scalable coding. Next, we show our proposed algorithm, and finally discuss results of experiment.

2. SNR Scalability

2.1 Decoder

In MPEG2 standard, base-layer information contains motion vectors, DCT coefficients and other important header information, and information of enhancement-layer contains residual DCT coefficients. Decoder uses the MV (Motion Vector) information taken from bit stream of base layer. Therefore, the motion vector should represent the motion of both layers. The decoding scheme of each layers is shown at Fig.1.



- Variable Length Decoding (VLD)
- Inverse Quantization (IQ)
- Inverse DCT (IDCT)
- Frame Store Memory (FS)
- Motion Compensation Prediction (MCP)

Fig 1. MPEG-2 Standard SNR Scalable Decoder

Both bit stream of base layer and enhancement layer are separately decoded and inverse-quantized. And then, DCT coefficients of both layers are added together.

We can say that DCT/IDCT operation is additive and Quantization / Inverse-Quantization is not additive. We do IDCT each layer first, and then add pixel value each other as shown at Fig.1.

Suppose that decoder first receives whole layer data. If bit stream of base, but not enhancement layer is received by decoder, 'drift' comes out. As number of processing frames increase, drift badly affects the PSNR of picture (PSNR value of picture will be reduced).

2.2 MPEG-2 compliant SNR scalable encoder structure

The 1-layer SNR scalable encoder produces SNR scalable video stream with simple structure.

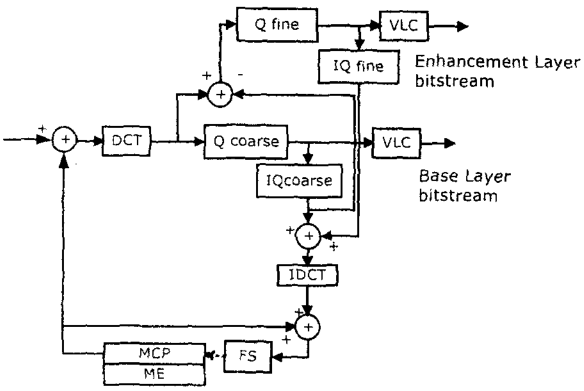


Fig 2. MPEG-Compliant 1-layer encoder

If decoder receive only bit stream of base layer, drift will occur. Because encoder reconstructs present picture including the information of enhancement layer. This encoder structure is compliant with MPEG-2 decoder.

The modified 1-layer is the drift-free 1-layer structure. This approach reconstructs present picture using only information of base layer. It also uses one motion vector representing the motion of both layers. It does not have drift, but have low coding efficiency and not MPEG-compliant.

Other approach is 2-Layer structure. MPEG-compliant 2-layer encoder is proposed to solve the drift problem and reduce coding rate. The key idea is to use separate encoding structure for each layers. The present 2-layer MPEG-compliant approach did not use enhancement information and still has coding efficiency problem.

3. Proposed Algorithm

3.1 Motivation

The present MPEG-compliant 2-layer algorithm uses only one MV of base layer. It is true that motion vector of base layer is optimum to base layer. But we code the enhancement layer by using MV of base layer. Our purpose is reducing the total bit rate of both layers by using one motion vector, which represents base layer and enhancement layer. The MV of enhancement layer is not well suitable for every picture of base layer. Now, we devise a linearly combined motion vector for base and enhancement layer. Detailed explanation of processes will be followed.

The concept of LCMV starts from present codec algorithm. The present MPEG-2 codec uses only MV of base layer. If we code the picture using information, which includes enhancement layer, total bit rate will be changed. Generally, when we use motion vector of enhancement layer, it will bring about increasing bit rate of base layer.

Therefore,, we need to find a scheme of LCMV which reduces total bit rate of base and enhancement layer.

We devise a expression of linearly combining two MVs as follows :

$$\mathbf{MV = aMVf + (1\text{-}a)MVb}$$

a is weight of enhancement layer.

The MVb is motion vector of base layer. But MVf is not vector of enhancement layer. The SNR scalable coding generates different SNR quality picture by using different quantization step size according to the layer. So, We set MVf as MV of full quality picture. (The prediction pictures are made of bit stream of base layer and enhancement layer.) Because the data of enhancement layer are generated from difference between present original picture and coarsely quantized picture (reconstructed frame of base layer.), we do not calculate the MV of enhancement layer directly.

3.2 Algorithm

We show our proposed encoder at Fig.3.

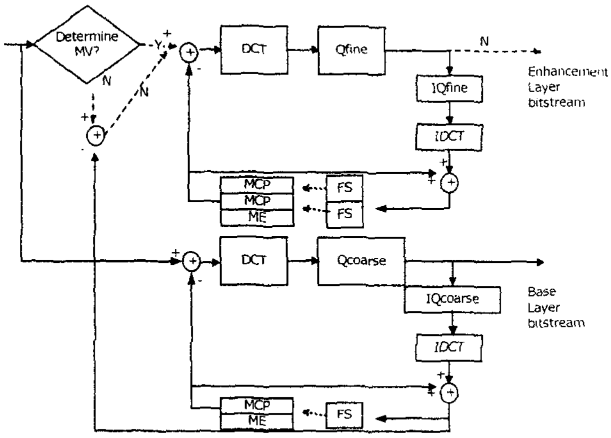


Fig 3. Proposed SNR scalable encoder

Algorithm of the proposed encoder is as follows.

1) start with initial values : a=0.5, b=0.5
2) obtain MVf and MVb
3) find delta(Δ) of Re & Rb for past two frames
4) calculate **a** of full quality layer information

$$a_i = a_{i-1} - \Delta a, \qquad 0 \le a_i \le 1$$

$$\Delta a = \frac{1}{c_1 2\pi}\arctan(c_2(\Delta R_e + \Delta R_b))$$

where, i : Frame index
c_1 and c_2 : Coefficients

5) Apply the above result to MV
6) MV will be transported via base-layer.

Where, MVf: MV of full quality picture.
MVb: MV of base layer.
Re : Rate of enhancement layer
Rb: Rate of base layer

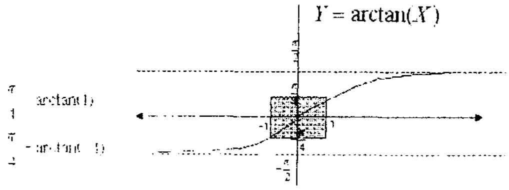


Fig 4.determination of Δa

We use arctangent function to decide delta of weight **a**. Arctangent function saturates the delta according to large increase and decrease of output of both layers. For avoiding severe fluctuation, we should damp the parameter of arctangent. We want to adjust the dynamic range of Y to be from -0.5 to 0.5. Therefore C1 =1/4. We assume that C2 is related to picture size (ex: QCIF, CIF) and complexity of sequence. In this experiment, we simply use 1/size as C2.

4. Experiments

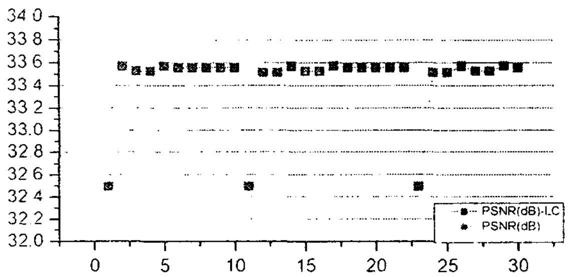
We use three sequences: sequence A is 'Mobile and Calendar', sequence B is 'Irene', and sequence C is 'Tempete'.

Picture size is CIF (352 x 288 pixel). The GOP structure is I PPP PPP PPP PP... (N =12, M=1: N; number of GOP, M: distance between I and P frame.). Each sequences consists of 30 frames. We compared the proposed algorithm with traditional two-layer algorithm. The quantization step size of base layer (Qb) is 20. The quantization step sizes of enhancement layer (Qf) are 5, 10, and 15.

The Simulation result is shown at Table.1.

Seq			Qb:20, Qf: 5	Qb:20, Qf: 10	Qb:20, Qf: 15
A	BPP	LCMV	2.3034	1.0182	0.5078
		Traditional	2.3017	1.0177	0.5071
	PSNR	LCMV	33.44	32.57	31.65
		Traditional	33.44	32.57	31.65
B	BPP	LCMV	0.7526	0.2745	0.1434
		Traditional	0.7768	0.2871	0.1500
	PSNR	LCMV	38.91	37.95	37.09
		Traditional	38.91	37.94	37.08
C	BPP	LCMV	1.7590	0.7048	0.3414
		Traditional	1.7755	0.7140	0.3459
	PSNR	LCMV	34.56	33.604	32.66
		Traditional	34.57	33.608	32.66

Table 1 Average bit rate and average PSNR over 30 frames



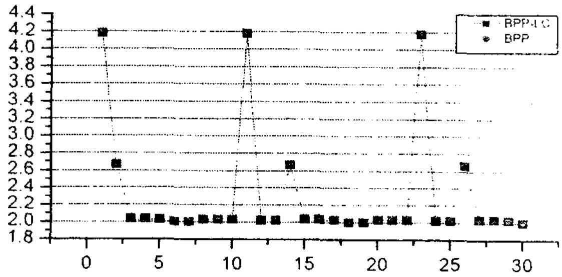


Fig 5 PSNR and BPP (Sequence A: Qb =20, Qf= 5)

Fig 6 Decoded Picture (Traditional / Linearly combining)

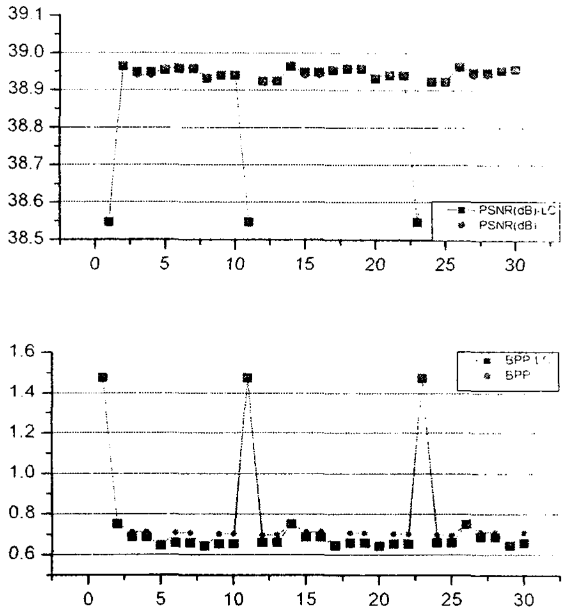


Fig 7. PSNR and BPP (Sequence B : Qb =20, Qf = 10)

Fig 8. Decoded Picture (Traditional / Linearly combining)

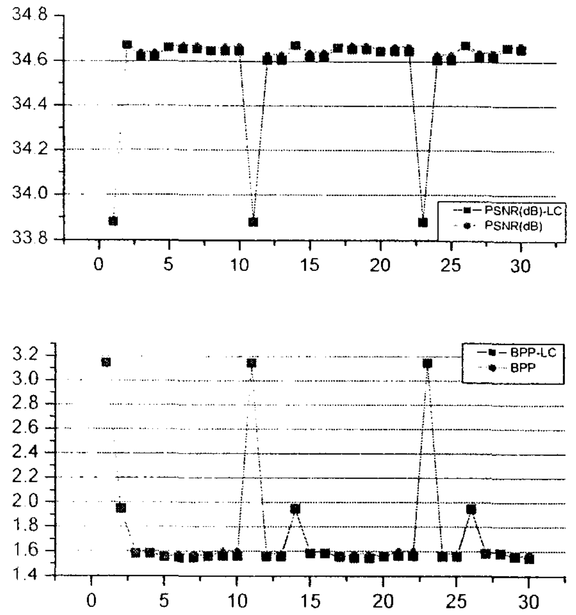


Fig 9. PSNR and BPP (Sequence C : Qb =20, Qf= 15)

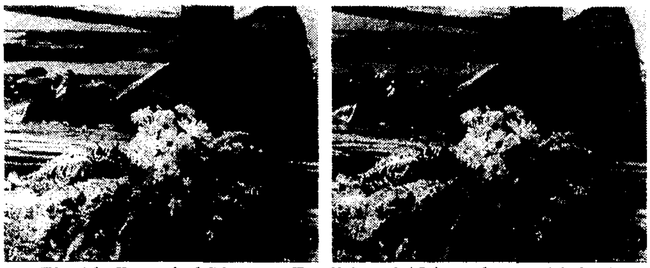


Fig 10. Decoded Picture (Traditional / Linearly combining)

Each sequences show same PSNR (linearly combining VS traditional). In view of bit rate, LCMV algorithm is almost same as traditional method in sequence A. In B sequence, LCMV is superior to traditional methods in bit rate. At the same PSNR, gains in bit rates are 3.3 %, 4.4%, and 4.5% in sequence B, and 1%, 2.3%, and 2.3% in sequence C respectively. Therefore, we can say that our scheme is effective in more complex sequences, but is ineffective in less complex sequence such as 'Mobile and Calendar Sequence'.

The result shows that the more it has complexity and motion, the less it has bit rate in simulation as shown in sequences B and C.

The proposed LCMV algorithm is not yet complete, and it needs too find proper coefficient C2 to give better bit rate for complex and fast sequence.

5. Conclusions

The SNR scalable coding has same spatial and temporal resolution, but has different SNR quality depending on network condition or user needs.

We proposed a new algorithm for MPEG-2 compliant SNR scalable encoder which has reduced coding rate. With proposed LCMV algorithm adapts well to complex and fast sequences in general.

References

[1] Arnold, "Efficient drift-free signal-to-noise ratio scalability.", IEEE 2000

[2] Kondi, "On video SNR scalability"

[3] Rose, "Toward Optimality in Scalable Predictive coding.", IEEE 2001

[4] Ghanbari, 'Video Coding', IEE Series, 2000

[5] Mitchell, 'MPEG Video Compression Standard', Chapman and Hall, 1996

[6] Sun, 'Compressed Video Over Networks', Marcel Dekker, 2001

[7] Haskell, 'Digital Video: An Introduction to MPEG-2', Chapman and Hall, 1997

[8] Bovik, 'Handbook of Image and Video Processing' Academic Press, 2000

[9] Gonzalez, 'Digital Image Processing', Addison Wesley, 1999

[10] Saywood, 'Introduction to Data Compression', Morgan Kaufmann, 2000

[11] Therrien, 'Discrete Random Signal and Statistical Signal Processing' 1992, Prentice-Hall

[12] R. Preuss, S. Roukos, A. Huggins, H. Gish, M. Bergamo, P. Peterson, "Embedded Signalling", *U. S. Patent* 5,319,735, 1994

[13] The Open Group, "Universal Unique Identifier", *CAE Specification*, DCE 1.1: Remote Procedure Call, Document Number C706, 1997

[14] J. Makhoul, "Linear Prediction: A Tutorial Review", *Proc. IEEE*, vol. 63, pp. 561~580, 1975

[15] A. V. Oppenheim, D. H. Johnson, "Discrete Representation of Signals", *Proc. IEEE*, vol. 60, No. 6, June 1972