Adaptive up-sampling method for H.264 scalable video coding

Ilhong Shin and HyunWook Park

Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
hwpark@athena.kaist.ac.kr

ABSTRACT

In a scalable video coding (SVC), spatial up-sampling of video sequences is an important process for spatial scalability. We propose an up-sampling method based on the type-1 DCT. For further improvement of the up-sampling method, we introduce an adaptive filtering method, which applies different weighting parameters to DCT coefficients. The proposed adaptive up-sampling method shows much improved PSNR in comparison with the recent H.264 SVC up-sampling method.

Index Terms— Adaptive filters, DCT, video codecs.

1. INTRODUCTION

Recently, H.264/AVC provides scalable video model for scalable extension of H.264/AVC. It inherited most building blocks of H.264 with some improved features for scalability such as UMCTF (unconstrained motion compensated temporal filtering) having no up-date step [1]. For spatial scalability, 11 taps and 6 taps filters were exploited in current JSVM (joint scalable video model) for down/up-sampling, respectively. Although their computational complexity and down and up-sampling performance are proper to be exploited in scalable video coding, more improvement must be made for rate-distortion performance of scalable video coding, *i.e.*, the performance of spatial scalability depends on the adopted down/up-sampling method.

Many researches were performed for image resizing by using DCT kernel employing characteristics of DCT [2-3]. However, this method can be applied only to the transcoding application.

We proposed a new spatial domain down/up-sampling method using type-2 DCT with half-pel shift in resizing [4]. When no-phase shift is expected like current coordinate of H.264 SVC, we propose type-1 DCT based up-sampling. In addition, we apply the adaptive filtering method in the type-1 DCT based up-sampling, which applies different weighting parameters to each DCT kernel. When this method is applied, up-sampling operation in the decoder has no difference in the computational complexity compared to the type-1 DCT

up-sampling with constant weighting parameters, where it can be called as the fixed type-1 DCT up-sampling.

In this paper, we introduce an fixed type-1 DCT up-sampling method in section 2. The proposed up-sampling method exploits various symmetries inherited from the type-1 DCT and IDCT kernel for fast algorithm, which is described in section 3. The adaptive up-sampling method is described in section 4. Experimental results and analysis of the proposed method are given in section 5, and section 6 concludes the paper.

2. Type-1 DCT up-sampling method

Since the symmetry assumption in type-1 DCT has no shift in spatial domain, the type-1 DCT has no phase-shift in down/up-sampling. Definition of non-orthogonal type-1 DCT is given as follows,

$$F(k) = \sqrt{\frac{2}{N}} \cdot \sum_{n_2=0}^N p(n_2) \cdot f(n_2) \cdot \cos\left(\frac{\pi k n_2}{N}\right) \quad (1)$$

$$\text{where } p(0) = p(N) = \frac{1}{2}, p(n_2) = 1$$

where $f(n_2)$ is the input signal having $N+1$ sample size. Figure 1 represents the up-sampling scheme using type-1 DCT. To make Nyquist interpolation kernel, the highest frequency signal in the transform coefficient should be halved. Then, N zeros are appended in the high frequency region. We can write the two-fold up-sampling process in a matrix form as follows:

$$\begin{aligned} B_{(2N+1) \times (2N+1)}^u &= \\ T_{(2N+1) \times (2N+1)}^t &\times \begin{bmatrix} T_{(N+1) \times (N+1)} \cdot B_{(N+1) \times (N+1)} \cdot T_{(N+1) \times (N+1)}^t & O_{(N+1) \times N} \\ O_{N \times (N+1)} & O_{N \times N} \end{bmatrix} \times T_{(2N+1) \times (2N+1)} \\ &= T_{(2N+1) \times (N+1)}^{L_t} \cdot T_{(N+1) \times (N+1)} \cdot B_{(N+1) \times (N+1)} \cdot T_{(N+1) \times (N+1)}^t \cdot T_{(2N+1) \times (N+1)}^L \\ &= V_{(2N+1) \times (N+1)}^U \times B_{(N+1) \times (N+1)} \times H_{(N+1) \times (2N+1)}^U \end{aligned} \quad (2)$$

where $T_{(N+1) \times (N+1)}$ denotes the 1D type-1 DCT kernel, and $B_{(N+1) \times (N+1)}$ and $B_{(2N+1) \times (2N+1)}^u$ are the original block and up-sampled block, respectively. $T_{(2N+1) \times (N+1)}^L$ represents left

the $(N+1)$ columns of $T_{(2N+1) \times (2N+1)}$, and the superscript t means transpose of the matrix. In eq. (2), $O_{N \times N}$ is the $N \times N$ zero matrix. The diagonal matrix for scaling is omitted for simplicity. $V_{(2N+1) \times (N+1)}^U$ and $H_{(N+1) \times (2N+1)}^U$ denote the vertical and horizontal up-sampling kernels, respectively. The transpose relationship of the vertical and horizontal up-sampling matrix holds as $V_{(2N+1) \times (N+1)}^{U,t} = H_{(N+1) \times (2N+1)}^U$. Therefore, two-fold type-1 DCT up-sampling matrix can be written as follows,

$$V^U(n_1, n_2) = \frac{2}{N} \sum_{k=0}^N s(k) \cdot p(n_2) \cdot \cos\left(\frac{\pi k n_2}{N}\right) \cdot \cos\left(\frac{\pi k n_1}{2N}\right)$$

where $s(0) = s(N) = p(0) = p(N) = \frac{1}{2}$, $s(k) = p(n_2) = 1$ (3)

for $1 \leq k \leq N-1$, $1 \leq n_2 \leq N-1$, $0 \leq n_1 \leq 2N$

The combined up-sampling kernel provides an efficient implementation structure. We employ an overlapping method to reduce blocking artifact and to improve the PSNR performance. In our implementation with the pixel overlapping, type-1 DCT with $N+1=13$ and type-1 IDCT with $2N+1=25$ samples are used for the two-fold up-sampling. The overlapped points after up-sampling are averaged to further reduce the blocking artifacts. Actually, for the sake of computational efficiency, the overlapping pixels after up-sampling can be simply truncated instead of averaging [5].

Fig. 1: Up-sampling method using type-1 DCT.

3. Fast algorithm of up-sampling

Since type-1 DCT up-sampling matrix follows Nyquist interpolation criteria, the even row of the up-sampling matrix can be removed in figuring. Therefore, $V_{N \times (N+1)}^{U,O}$ represents odd row up-sampling kernel. We can show the up-sampling matrix of type-1 DCT is symmetric, i.e., $V_{N \times (N+1)}^{U,O}(n_1, n_2) = V_{N \times (N+1)}^{U,O}(N-1-n_1, N-n_2)$. This symmetry can be exploited for reducing the number of multiplications of the up-sampling kernel like type-2 DCT up-sampling [4]. Since the column length of the type-1 DCT based up-sampling is odd, We can change the odd column-size up-sampling matrix to even column-size matrix as follows,

$$\begin{bmatrix} a & e & d \\ b & f & c \\ c & f & b \\ d & e & a \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a & \frac{e}{2} & \frac{e}{2} & d \\ b & \frac{f}{2} & \frac{f}{2} & c \\ c & \frac{f}{2} & \frac{f}{2} & b \\ d & \frac{e}{2} & \frac{e}{2} & a \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_2 \\ x_3 \end{bmatrix} \quad (4)$$

We define F_1, F_2, B_1 , and B_2 as follows,

$$B_1(n_1, n_2) = B_{(N+2) \times (N+1)}^{U,O}(n_1, n_2) - B_{(N+2) \times (N+1)}^{U,O}(N+1-n_1, n_2)$$

$$B_2(n_1, n_2) = B_{(N+2) \times (N+1)}^{U,O}(n_1, n_2) + B_{(N+2) \times (N+1)}^{U,O}(N+1-n_1, n_2)$$

$$F_1(n_1, n_2) = V_{N \times (N+2)}^{U,O}(n_1, n_2) - V_{N \times (N+2)}^{U,O}(n_1, N-n_2+1)$$

$$= \sum_{k=0}^N s(k) \cdot p'(n_2) \cdot \left(\frac{1-(-1)^k}{2}\right) \cdot \cos\left(\frac{\pi k n_2}{N}\right) \cdot \cos\left(\frac{\pi k (2n_1+1)}{2N}\right) \quad (5)$$

$$F_2(n_1, n_2) = V_{N \times (N+2)}^{U,O}(n_1, n_2) + V_{N \times (N+2)}^{U,O}(n_1, N-n_2+1)$$

$$= \sum_{k=0}^N s(k) \cdot p'(n_2) \cdot \left(\frac{1+(-1)^k}{2}\right) \cdot \cos\left(\frac{\pi k n_2}{N}\right) \cdot \cos\left(\frac{\pi k (2n_1+1)}{2N}\right)$$

$$p'(n_2) = \begin{cases} \frac{1}{2}, & n_2 = 0, \frac{N}{2} \\ 1, & \text{otherwise} \end{cases} \text{ for } 0 \leq n_1 < \frac{N}{2}, 0 \leq n_2 < \frac{N}{2} + 1$$

Therefore, vertical filtering of odd index points using eqs. (5) can be re-written as follows,

$$V_{N \times (N+1)}^{U,O} \times B_{(N+1) \times (N+1)} = V_{N \times (N+2)}^{U,O} \times B_{(N+2) \times (N+1)} =$$

$$\begin{bmatrix} F_{1, \frac{N}{2} \times (\frac{N}{2}+1)} \times B_{1, (\frac{N}{2}+1) \times (N+2)} + F_{2, \frac{N}{2} \times (\frac{N}{2}+1)} \times B_{2, (\frac{N}{2}+1) \times (N+2)} \\ F_{1, \frac{N}{2} \times (\frac{N}{2}+1)}^r \times B_{1, (\frac{N}{2}+1) \times (N+2)} - F_{2, \frac{N}{2} \times (\frac{N}{2}+1)}^r \times B_{2, (\frac{N}{2}+1) \times (N+2)} \end{bmatrix} \quad (6)$$

where r denotes the row inversion, which changes each column elements to the reverse direction. The horizontal up-sampling can be performed in a similar way. Further reduction is still possible with the representation of F_2 filter as $F_2 = V_{N/2 \times (N/2+1)}^{U,O}$. It shows F_2 filter contains type-1 DCT up-sampling filter with a half of N . The decomposition of F_2 filter can be possible until F_2 filter becomes 2×2 with some manipulations.

Table 1 shows the up-sampling complexity of the H.264 SVC, the proposed up-sampling without fast algorithm, and the proposed up-sampling with fast algorithm.

Table 1. The complexity in two-fold up-sampling (N=12 was applied, 8×8 block is up-sampled to 16×16)

Up-sampling method	Multiplication
Proposed without fast method	3016
Proposed with fast method	986
H.264 SVC	584

4. Adaptive up-sampling method

We extend our up-sampling method by using adaptive filtering scheme. We employ the type-1 DCT up-sampling matrix in the adaptive filtering, which determines relative weighting of each frequency component of type-1 DCT. For adaptive up-sampling, we derive a vertical adaptive up-sampling matrix as follows,

$$V_{(N) \times (N+1)}^{U,O,A} = T_{(N) \times (N+1)}^{L,O,I} \cdot P_v \cdot T_{(N+1) \times (N+1)} \quad (7)$$

$$\text{where } P_v = \begin{pmatrix} v_0 & 0 & \dots & 0 \\ 0 & v_1 & 0 & \dots & 0 \\ 0 & 0 & v_2 & 0 \dots & 0 \\ 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & v_N \end{pmatrix}$$

where O denotes odd row of matrix. The proposed adaptive up-sampling matrix with the weighting parameters can be made as follows,

$$V_{(N) \times (N+1)}^{U,O,A}(n_1, n_2) = \frac{2}{N} \sum_{k=0}^{N-1} s(k) \cdot p(n_2) \cdot p_v(k) \cdot \cos\left(\frac{\pi k n_2}{N}\right) \cdot \cos\left(\frac{\pi k}{N} \left(n_1 + \frac{1}{2}\right)\right) \quad (8)$$

$$\text{where } s(0) = s(N) = p(0) = p(N) = \frac{1}{2},$$

$$s(k) = p(n_2) = 1 \text{ for } 1 \leq k \leq N-1, 1 \leq n_2 \leq N-1, 0 \leq n_1 \leq N-1$$

The adaptive weighting parameters are inserted into eq. (3), and only odd row up-sampling kernel is made. It can be understood that the adaptive up-sampling adjusts the cosine basis function of forward and inverse DCT in consideration with the image characteristics. The required

number of adaptive parameters is $N-1$. Since v_0 controls DC component, it always set to 1., and v_n is 0 in eq. (8).

When the proposed adaptive up-sampling method is applied, the up-sampling matrices of $V^{U,O,A}$ and $H^{U,O,A}$ are computed and transmitted for every frame by using the optimum weighting parameters of P_v and P_h , respectively. The entropy coding of adaptive parameters were employed with UVLC [1]. The computational complexity of up-sampling matrix is same as that of all weighting parameters of one, which is the fixed type-1 DCT up-sampling matrix. Therefore, we can apply the proposed fast algorithm of previous section with suitable modification.

We used least-square minimization method for obtaining adaptive parameters. Figure 2 shows the sample position in two-fold up-sampling.

Fig. 2: Sample positions of two-fold up-sampling.

We obtain the optimum P_v that minimizes error between the original up-sampled image and the computed up-sampled image as follows,

$$\sum_{i=1}^{NB} \left| B_{orig}^{i,c} - T_{(N) \times (N+1)}^{L,O,I} \cdot P_v \cdot T_{(N+1) \times (N+1)} \cdot B_{(N+1) \times (N+1)}^i \right|^2 \quad (9)$$

where $B_{orig}^{i,c}$ is the i 'th block of the original up-sampled image, which are set of pixels such as position 'c' in Fig. 2, and NB is the number of blocks in the down-sampled image. As we can see in eq. (9), the diagonal matrix P_v can not be represented by solvable form due to nature of matrix multiplication. We should translate eq. (9) as solvable form to avoid mathematical difficulty with kronecker tensor product as follows,

$$\left| B_{orig}^{c,s} - K^v \cdot \text{Diag}(P_v) \right|^2 \text{ where } B_{orig}^{c,s} = \begin{pmatrix} B_{N \times (N+1)}^{1,c,s} \\ \vdots \\ B_{N \times (N+1)}^{NB,c,s} \end{pmatrix} \text{ and } K^v = \begin{pmatrix} R_{N \times N}^{1,v} \\ \vdots \\ R_{N \times N}^{NB,v} \end{pmatrix} \quad (10)$$

$$R_{N \times N}^{i,v} = [A_1^{i,v} \ A_{N+2}^{i,v} \ \dots \ A_{N^2+N}^{i,v}], A^{i,v} = ((B_{(N+1) \times (N+1)}^{i,t} \cdot T_{(N+1) \times (N+1)}^t) \otimes T_{(N) \times (N+1)}^{L,O,I})$$

where s and subscript denote stacked operator and column indices of matrix, respectively. We omit complex derivation process for simplicity. Therefore, we can obtain optimal solution as follows,

$$\text{Diag}(P_v) = (K^{v,t} \cdot K^v)^{-1} \cdot K^v \cdot (B_{orig}^{c,s}) \quad (11)$$

Also we can obtain horizontal minimization form as follows,

$$\left| B_{orig}^{b,s} - K^h \cdot \text{Diag}(P_h) \right|^2 + \left| B_{orig}^{d,s} - K^d \cdot \text{Diag}(P_h) \right|^2 \quad (12)$$

The horizontal parameters should consider the diagonally up-sampled position. The optimal solution can be written as follows,

$$\begin{aligned} \text{Diag}(P_h) \\ = (K_r^{h,t} \cdot K_r^h + K_r^{d,t} \cdot K_r^d)^{-1} \cdot (K_r^{h,t} \cdot B_{orig}^{b,s} + K_r^{d,t} \cdot B_{orig}^{d,s}) \end{aligned} \quad (13)$$

5. Experimental result

We implemented the proposed adaptive type-1 DCT up-sampling method in the recent JSVM video codec [1]. We used 4CIF video sequences of City and Harbour for experimental evaluation. The base layer was encoded with 352×288 (CIF) in the 4CIF sequence after down-sampling by JSVM method. The proposed adaptive up-sampling method will be efficient for high quality base-layer, *i.e.*, less quantization of base layer images gives more accurate up-sampling results. However, very low QP in the base-layer may degrade the coding performance. Therefore, we used a moderate experimental condition, where QP of base-layer is set to 24 and 30. All frames were coded as intra frame to show maximum performance of the proposed method.

We can see the proposed adaptive up-sampling method improves the PSNR performance in Fig. 3 with high quality base-layer coding using QP of 24. The maximum improvements of PSNR are about 0.4 dB (a) and 0.65 dB (b). The fixed up-sampling denotes the proposed type-1 DCT up-sampling without adaptive parameter. We can see the adaptive up-sampling method improves 0.2~0.3 dB compared to the fixed up-sampling method. The improvements of PSNR are 0.25 dB (a) and 0.35 dB (b) in the middle quality base layer coding with QP 30. The proposed adaptive up-sampling method still provides more improved result with middle quality coding. However, H.264 SVC provides any condition of base layer video quality, so the proposed adaptive up-sampling method will be efficient at the high and middle quality of base-layer for spatial scalability.

6. Conclusions

We proposed an adaptive up-sampling method for performance improvement of spatial scalability in the H.264 SVC. We developed the up-sampling method using type-1 DCT without phase shift. Also we proposed a fast algorithm for up-sampling using symmetries of DCT kernel. By transmitting the adaptive parameters of type-1 DCT up-sampling kernel, we showed the improved result of the proposed adaptive up-sampling method in

comparison with the JSVM up-sampling method under the proper assumption of base layer quality. We are expecting that the proposed method can improve the spatial scalability performance of H.264 SVC.

7. REFERENCES

- [1] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 'Joint Scalable Video Model JSVM-3', JVT-Q202, Nice, France, Oct. 2005.
- [2] R. Dugad and N. Ahuja, "A fast scheme for image size change in the compressed domain," IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 461-474, Apr. 2001.
- [3] H. W. Park, Y. S. Park, and S. K. Oh, "L/M fold image resizing in block-DCT domain using symmetric convolution," IEEE Trans. Image Processing, vol. 12, pp. 1016-1034, Sept. 2003.
- [4] I.H. Shin and H.W. Park, "Efficient down-up sampling using DCT kernel for MPEG-21 SVC," in Proc. Int. Conf. Image Processing, 640-643 Sept. 2005.
- [5] I.H. Shin and H.W. Park, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 'Down/up-sampling method for spatial scalability of SVC', JVT-0024, Busan, Korea, Apr. 2005.

Fig. 3. Rate-PSNR curves of the JSVM and proposed method for (a) City and (b) Harbour.