A fast down-sampling method for arbitrary-ratio spatial scalability based on type-II DCT

Ilhong Shin, Jeongju Yu,and JinWoo Hong
Electronics Telecommunications Research Institute, Daejeon, South Korea
ssi@etri.re.kr

ABSTRACT

In a scalable video coding (SVC), spatial down/up-sampling of video sequences is an essential process for spatial scalability. We propose an arbitrary-ratio spatial down-sampling method based on the type-II DCT, with which down-sampling of H.264 SVC can be efficient. Simple windowing operation in the DCT domain is efficiently embedded in the proposed down-sampling method for aliasing reduction. A fast algorithm for arbitrary-ratio down-sampling is introduced with matrix decomposition and fast DCT. It achieves 80 % reduction of computation complexity in comparison with the direct matrix calculation.

Index Terms— Adaptive filters, DCT, image processing, video codecs, Z transform.

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 hierarchical B pictures [1]. For spatial scalability, 12 taps and 4 taps filters were exploited in current JSVM (joint scalable video model) for down/up-sampling, respectively. Although current down-sampling method provides good visual quality, the complexity is burden to real-time encoding system.

We proposed a down/up-sampling method for only two-fold case with type-II DCT [2]. However, current standard should provide arbitrary-ratio up-sampling with phase shift in down/up-sampling. Previous fast algorithm [2] can not support the arbitrary-ratio up-sampling efficiently, because it does not exploit the characteristics of DCT. Furthermore, only discarding high frequency component in the DCT domain [3-7] make aliased down-sampled image, where visual quality is important in the image or video down-sampling process.

In this paper, we introduce an arbitrary-ratio spatial down-sampling method based on type-II DCT in section 2, where frequency response of proposed method is well fitted to that of JSVM. The proposed down-sampling method exploits various symmetries inherited from the type-II and IV DCT and IDCT kernels for fast algorithm, which is described in section 3, and section 4 concludes the paper.

2. Type-II DCT down-sampling method

The current H.264 SVC uses down/up-sampling with various ratios, which has phase shift. Since the up-sampling coordinate should be matched to the down-sampling coordinate, the proposed down-sampling filter should have a corresponding phase shift. Since H.264 SVC provides arbitrary-ratio spatial scalability, previous approach [2] should be generalized for arbitrary-ratio spatial up-sampling. We generally define the down-sampling ratio as L/M, where L and M can be any integer numbers with L<M. As the type-II DCT has a phase shift in down/up-sampling, the down/up-sampling using type-II DCT has a corresponding coordinate of H.264 SVC. When the signal length of the type-II DCT is M, M-L zeros are discarded in the high frequency region after the type-II DCT. Following this, the type-II IDCT of the shortened L samples is performed to obtain L/M-fold down-sampled data. As the cascaded operation in down-sampling using DCT and IDCT is inefficient in terms of computational complexity, a combined operation of type-II DCT down-sampling is proposed. The L/M-fold down-sampling process can be described in a matrix form as follows:

$$B_L^d = \begin{bmatrix} T_L^t \cdot P & O_{(M-L)} \end{bmatrix} \times T_M \cdot B_M = V_{L\times M} \cdot B_M \tag{1}$$

where T_M denotes the 1D type-II DCT kernel for M samples, and B_M and B_L^d are the original image block of $M\times M$ and the down-sampled image block of $L\times L$, respectively. T_L^t represents the type-III DCT of L samples, and the superscript t indicates the transpose of the matrix. The inverse transform of the type-II DCT is simply type-III DCT. In Eq. (1), $O_{(M-L)}$ is the zero matrix of $(M-L)\times(M-L)$ size, which is required for truncation of high frequency component in the DCT domain, and P represents diagonal matrix for the adjustment of DCT coefficient, where it is efficiently used with strong anti-aliasing. $V_{L\times M}$ is the combined down-sampling matrix for vertical down-sampling. The scaling factor in the matrix calculation is omitted for simplicity. It should be noted that one-dimensional process of proposed method is easy to make the two-dimensional down-sampling. The horizontal down-sampling matrix is simply transpose of that of vertical one. The previous down-sampling method with DCT [2] set P matrix as identity. The frequency response of JVT filter is shown at Fig. 1 (Method 1), where it shows strong anti-aliasing with sacrificing detail preservation. However, we adopt JSVM filter for visual appearance. Proposed one finds optimal weighting parameter P for having similar frequency response of JVT filter. We used least square optimization method for determining P matrix. In other word, we searched optimal P matrix with frequency response of JVT filter. The frequency response of DCT based down-sizing is written as follows:

$$D(z) = \frac{1}{M}\sum_{k=0}^{M-1} B(e^{-\frac{j\cdot 2\cdot\pi\cdot k}{M}}\cdot z^{\frac{L}{M}})\cdot F_k(e^{-\frac{j\cdot 2\cdot\pi\cdot k}{M}}\cdot z^{\frac{L}{M}}) \tag{2}$$

$$F_k(z) = \sum_{i=0}^{L-1} z^{-\frac{M\cdot i}{L}} \cdot H_{DCT,L/M,P,i}(z)\cdot e^{-\frac{j\cdot 2\cdot\pi\cdot k}{L}} \tag{3}$$

where $H_{DCT,L/M,P,i}(z)$ is a z-transform of M-tap filter,

This work was supported by Ministry of Information and Communication (MIC) of Korean government (2005-S-103-03).

which is represented by ith row of $V_{L\times M}$ matrix. As shown in [5], since the magnitude of $F_0(z)$ is dominant in comparison with the other component, we only deal with the frequency response of $F_0(z)$ for deriving proposed filter. The problem of finding optimal P matrix is written as follows:

$$\arg\underset{P}{Min}\ (|H_{JVT}(z)|-|F_0(z)|)^2 \tag{4}$$

where $|\cdot|$ denotes the magnitude of z-transformed result, and $H_{JVT}(z)$ is the z-transform of JVT's down-sampling filter. However, direct calculation of P matrix is impossible due to the non-linear nature of problem. We used Levenverg-Marquardt optimization method for finding P matrix. We used L/M as 8/16 for two-fold down-sampling, which is convenient for clarification. The obtained P matrix is written as follows:

$$diag(P)= \{1, 1.0048, 1.0048, 1.0208, 1.0200, 0.8080, 0.6288, 0.0624\} \tag{5}$$

where $diag(\cdot)$ denotes the diagonal elements of matrix. The obtained weighting parameters decreased at high frequency index, hence it reduces the aliasing caused by high frequency data, while lessening detail of image.

Figure 1 (a) shows the frequency response of JVT filter (method 1), previous method (method 2), and proposed method. Method 1 shows strong anti-aliasing, whereas method 2 shows severe aliasing due to the sharp transition. The frequency response of proposed method shows similar shape of method 1. However, attenuation at the high frequency band is shown, but the visual appearance shows similar result. Large number of M may improves the frequency response with increased complexity. Figure 1 (b) shows frequency response of $F_k(z)$, which includes aliasing response of multi-rate system applied to the down-sampling with DCT. However, figure 1 (c) shows reduced response in comparison with Fig 1 (b), because strong anti-aliasing is applied with proposed P matrix.

The visual appearance of "Mobile and Calendar" is shown in Fig 2. Fig. 2 (a) shows good compromise with reduced aliasing and lessening details of image. But, Fig. 2 (b) shows severe aliasing with previous method. The visual appearance of the proposed method is similar to that of JVT in Fig. 2 (c). However, small aliasing effect is still remained.

Table 1 shows possible values of L and M in the L/M-fold arbitrary-ratio down-sampling, where noted points were selected suitably in consideration with performance and complexity.

Table 1. Possible data points of forward and backward DCT for down-sampling ratio as L/M.

down-sampling ratio	Forward and backward DCT	
	M-points DCT	L-points IDCT
3/4	16	12
2/3	12	8
1/2	16	8

3. Fast algorithm of arbitrary-ratio spatial down-sampling

Each element of the L/M-fold type-II DCT-based down-sampling matrix of the vertical direction can be written as follows:

$$v_{L\times M}(n_1,n_2)=\frac{2}{M}\cdot\sum_{k=0}^{L-1}P_k\cdot c(k)\cdot\cos\left(\frac{\pi k(2n_2+1)}{2M}\right)\cdot\cos\left(\frac{\pi k(2n_1+1)}{2L}\right) \tag{6}$$

$$where\ \ c(k)=\frac{1}{2},\ c(k)=1\ \ for\ \ 1\le k\le L-1$$

where $v_{L\times M}(n_1,n_2)$ is the (n_1 , n_2)-element of the vertical downpling matrix, $V_{L\times M}$, and P_k is the k-th diagonal element of P matrix. The combined kernel of the forward and backward DCT has a large degree of symmetry. From Eq. (6), it is shown that the down-sampling kernel is symmetric, *i.e.*, $v_{L\times M}(n_1,n_2)=v_{L\times M}(L-n_1-1,M-n_2-1)$. In order to reduce the number of multiplications of the down-sampling kernel, the symmetric property can be exploited, and we can define two matrices of V_1 and V_2, whose (n_1, n_2)-elements are given as follows:

$$\begin{aligned}
&v_{1,\frac{L}{2}\times\frac{M}{2}}(n_1,n_2)=(v_{L\times M}(n_1,n_2)-v_{L\times M}(n_1,M-n_2-1))/2\\
&=\frac{2}{M}\cdot\sum_{k=0}^{\frac{L}{2}-1}P_{2k+1}\cdot c(k)\cdot\cos\left(\frac{\pi(2k+1)(2n_2+1)}{2\cdot M}\right)\cdot\cos\left(\frac{\pi(2k+1)(2n_1+1)}{2\cdot L}\right)\\
&v_{2,\frac{L}{2}\times\frac{M}{2}}(n_1,n_2)=(v_{L\times M}(n_1,n_2)+v_{L\times M}(n_1,M-n_2-1))/2\\
&=\frac{2}{M}\cdot\sum_{k=0}^{\frac{L}{2}-1}P_{2k}\cdot c(k)\cdot\cos\left(\frac{\pi k(2n_2+1)}{2\cdot M}\right)\cdot\cos\left(\frac{\pi k(2n_1+1)}{2\cdot L}\right)\\
&where\ \ c(k)=\frac{1}{2},\ c(k)=1\ \ for\ \ 1\le k\le L-1
\end{aligned} \tag{7}$$

Then, it is possible to decompose the vertical down-sampling matrix as follows:

$$\begin{aligned}
&V_{L\times M}\times B_M=V\times\begin{bmatrix} I_{M/2} & I_{M/2}\\ -R_{M/2} & R_{M/2}\end{bmatrix}\times\begin{bmatrix} S_M\\ A_M\end{bmatrix}\times B\\
&=\begin{bmatrix} V_1\cdot S_M+V_2\cdot A_M\\ R_{L/2}\cdot(V_2\cdot A_M-V_1\cdot S_M)\end{bmatrix}\times B\\
&where\ S_M=[I_{M/2}\ \ \text{-}R_{M/2}],\ A_N=[I_{M/2}\ R_{M/2}],\ R_M=\begin{bmatrix}0\ldots 0\,1\\ \vdots\ \ddots\ \vdots\\ 1\ \ldots 0\,0\end{bmatrix}.
\end{aligned} \tag{8}$$

In Eq. (8), $I_{M/2}$ and $R_{L/2}$ is the identity matrix and anti-identity matrix, respectively. V_2 is also symmetrical in Eq. (7), *i.e.*, $V_2(\frac{L}{2}-n_1-1,\frac{M}{2}-n_2-1)=V_2(n_1,n_2)$. By using the symmetry of V_2, further decomposition is possible, where V_2 is decomposed into $V_{1,\frac{L}{4}\times\frac{M}{4}}$ and $V_{2,\frac{L}{4}\times\frac{M}{4}}$ through a similar manner as that given by Eq. (7). Therefore, V_2 can be recursively decomposed until either L or M becomes an odd length. We can represent Eq. (8) in another form as follows:

$$\begin{aligned}
V_{1,\frac{L}{2}\times\frac{M}{2}}&=\left[C_{L/2}^{IV}\cdot P_o\ \ O_{(M-L)/2}\right]\cdot\left[C_{M/2}^{IV}\right]\\
V_{2,\frac{L}{2}\times\frac{M}{2}}&=\left[C_{L/2}^{III}\cdot P_e\ \ O_{(M-L)/2}\right]\cdot\left[C_{M/2}^{II}\right]
\end{aligned} \tag{9}$$

This work was supported by Ministry of Information and Communication (MIC) of Korean government (2005-S-103-03).

where the matrix of $C_{L/2}^{IV}$ denotes the type-IV DCT kernel. P_e and P_o denote even and odd rows of diagonal P matrix, respectively. Therefore, the decomposed matrix of the up-sampling kernel consists of the cascaded computation of $M/2$-points type-II DCT followed by $L/2$-points type-III DCT and $M/2$-points type-IV DCT followed by $L/2$-points type-IV DCT with proper truncation of high frequency component. When the fast DCT algorithm of each type is applied, the number of multiplications can be further reduced. However, direct cascaded implementation of type-IV DCT shows some inefficiency due to the nature of the fast type-IV DCT and figuring with P matrix in the DCT domain. We also propose a very efficient form of the cascaded implementation of V_1 filtering using the fast type-IV DCT [4-5]. The V_1 filtering can be written as follows:

$$Y = \left[C_{L/2}^{IV} \cdot P_e \;\; O_{(M-L)/2} \right] \cdot \left[C_{M/2}^{IV} \right] \cdot X \tag{10}$$

where X is the 1-D $M/2$-point input signal, and Y is the filtered output signal, whose size is $L/2$. Two fast algorithms for type-IV DCT were proposed by Chan [8] and Lee [9]. The number of multiplications for V_1 filtering using Chan or Lee's algorithms can be given as follows:

$$\begin{aligned} Mult(V_1) &= Mult(C_{L/2}^{IV}) + Mult(C_{M/2}^{IV}) \\ &= Mult(C_{L/2}^{II}) + Mult(C_{M/2}^{II}) + L \end{aligned} \tag{11}$$

The proposed combined approach uses Lee's algorithm prior to Chan's algorithm for V_1 filtering as follows:

$$\begin{aligned} & y(n) + y(n-1) \\ &= \sum_{k=0}^{\frac{L}{2}-1} [P_{e,k} \cdot \cos(\frac{2\pi(2k+1)}{4L}) \cdot \sec(\frac{2\pi(2k+1)}{4M}) \cdot X_C^{III}(k)] \cdot \cos(\frac{2\pi n(2k+1)}{2L}) \end{aligned} \tag{12}$$

where $\frac{M}{2}$-points type-III DCT of X_C^{III} is performed on $x(n)+x(n-1)$ instead of $x(n)$, as shown in Eq. (10). Therefore, the V_1 filtering can be performed by sequential operation of Lee's fast type-IV DCT followed by Chan's fast type-IV IDCT. The detail derivation is omitted due to page length of this paper. Since $P_{e,k} \cdot \cos(\frac{2\pi(2k+1)}{4L}) \cdot \sec(\frac{2\pi(2k+1)}{4M})$ in Eq. (12) can be pre-computed, the number of multiplications for V_1 filtering is finally reduced to:

$$Mult(V_1) = Mult(C_{L/2}^{II}) + Mult(C_{M/2}^{III}) + \frac{L}{2} \tag{13}$$

Eq. (13) has a smaller number of multiplications than Eq. (10). In addition, the total number of multiplications in $V_{L\times M}$ filtering can be expressed as follows:

$$\begin{aligned} Mult(V_{L\times M}) &= Mult(C^{III}_{L/2^r}) + Mult(C^{II}_{M/2^r}) \\ &+ \sum_{i=1}^{r} \left(Mult(C^{II}_{L/2^i}) + Mult(C^{III}_{M/2^i}) + L/2^i \right) \end{aligned} \tag{14}$$

where r implies the final stage of matrix decomposition. The proposed fast filtering method reduces the number of multiplications of the proposed up-sampling by approximately 80 % in comparison with the direct matrix calculation, as shown in Table 2. Also the proposed method reduces the computation complexity by 35 % in comparison with our previous method [2], where it is only for two-fold down-sampling. Furthermore, the proposed method has very low complexity about the H.264 SVC 12-tap down-sampling filter. Since the proposed method is based on block by block operation, the parallel operation is easy for further speed-up.

This work was supported by Ministry of Information and Communication (MIC) of Korean government (2005-S-103-03).

Table 2. Comparison of computational complexity.

L/M	Multiplication per pixel		
	H.264 SVC	Previous[2]	Proposed
3/4	15.75	Unavailable	3.28
2/3	13.3	Unavailable	2.08
1/2	9.0	3.18	2.34

4. Conclusions

We proposed simple and efficient arbitrary-ratio down-sampling method for spatial scalability of H.264 SVC. The experimental result shows that proposed one improves visual quality while reducing aliasing artifact. The adjusting of frequency response in the DCT-domain shows similar objective and subjective effects about JSVM filter. Also we proposed a fast algorithm for down-sampling using symmetries of DCT kernel and fast type-II and IV DCT, wherein the efficient method was developed by reducing redundant operation of cascaded type-IV DCT. The proposed method has low computational complexity in comparison with one of JVT.

5. REFERENCES

[1] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 'Joint Scalable Video Model JSVM-4', JVT-T202, Klagenfurt, Austria, Jul. 2006.

[2] 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.

[3] J. Mukherjee and S. Mitra, "Image resizing in the compressed domain using subband DCT," *IEEE Trans. Circuits Syst. Video Technol*., vol. 12, pp. 620-627, July. 2002.

[4] 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.

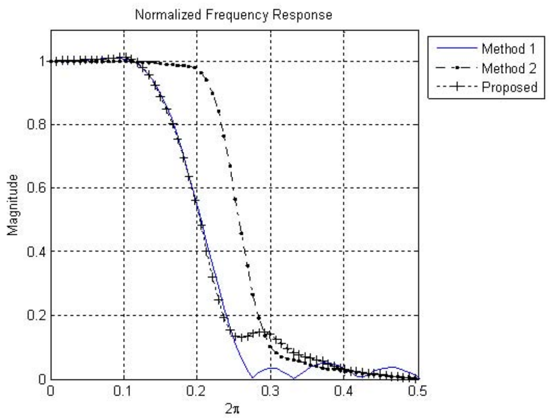
[5] C. L. Salazar and T. D. Tran, "On resizing images in the DCT domain," *in Proc. Int. Conf. Image Processing,* Oct. 2004.

[6] T. Frajka,and K Jegger, "Downsampling dependent upsampling of images," *Signal Processing : Image Communication,* vol. 19, pp. 257-265, March. 2004.

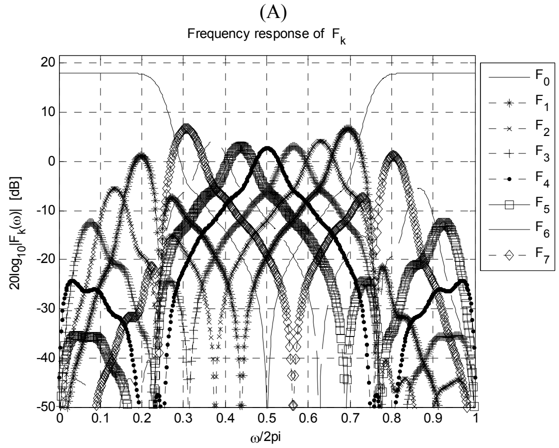
[7] Y. S. Park and H. W. Park, "Design and analysis of image resizing filter in the block-DCT domain," *IEEE Trans. Circuits Syst. Video Technol*., vol. 14, no. 2, pp. 274-279, Feb. 2004.

[8] S. C. Chan and K. L. Ho, "Direct method for computing the discrete sinusoidal transforms," *Proc. IEE. Part F,* vol. 137, pp. 433-442, Dec. 1990.

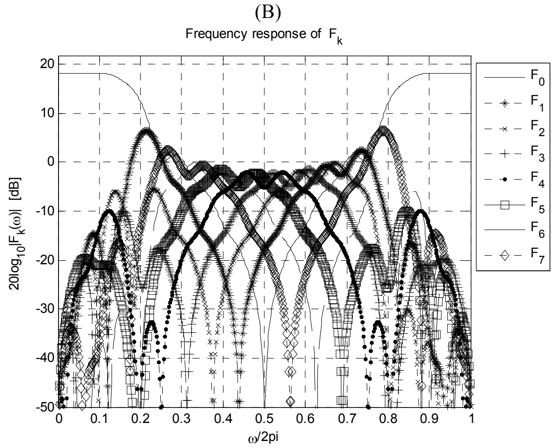
[9] B. Lee, "A new algorithm to compute discrete cosine transform," *IEEE Trans. ASSP*, vol. 32, no. 6, pp. 1243-1245, Dec. 1984.



(A)



(B)



(C)

Fig. 1: Frequency response of various filters: (a) Normalized response between JVT, previous method, and proposed one, (b) frequency response of $F_k(z)$ with P as identity, and (c) frequency response of $F_k(z)$ with proposed P.

(A)

(B)

(C)

Fig. 2: The down-sized frame of Mobile Calendar sequence : (a) JVT filter (QCIF), (b) previous method (QCIF), and (c) propose method (QCIF).

This work was supported by Ministry of Information and Communication (MIC) of Korean government (2005-S-103-03).