Adaptive Up-Sampling Method Using DCT for Spatial Scalability of Scalable Video Coding

IlHong Shin and Hyun Wook Park, *Senior Member, IEEE*

***Abstract*—In the incoming amendment of H.264/AVC for scalable video coding (SVC), down and up-samplings in a spatial domain are incorporated for spatial scalability. A new two-fold up-sampling method is proposed for SVC. The proposed up-sampling method is based on a combination of the forward and backward type-II discrete cosine transform (DCT). As the proposed up-sampling kernel has various symmetries for efficient computation, a fast algorithm of type-II DCT-based up-sampling method is also proposed. For further improvement of the up-sampling performance, an adaptive filtering method in the type-II DCT up-sampling is introduced, which applies different weighting parameters to DCT coefficients. Even as this adaptive method is applied, the up-sampling operation in the decoder has the same computational complexity as the fixed weighting method. The proposed adaptive up-sampling method shows a much improved PSNR in comparison with the recent H.264 SVC up-sampling method.**

***Index Terms*—Adaptive up-sampling, down-sampling, H.264, scalable video coding (SVC), type-II DCT, up-sampling.**

I. INTRODUCTION

THE EMERGING home-network and broadband convergence network technologies allow people to see images and video anywhere at anytime. The display devices can be a large plasma display panel, a standard definition TV, or a very small cell-phone. Therefore, the video codec used in these devices should support various displays with different resolutions.

Transcoding [1]–[3] can be a solution for the support of various video resolutions. a large amount of research has centered on image resizing using a discrete cosine transform (DCT) kernel and employing the characteristics of DCT [4]–[6]. Down/up-sampling using the DCT kernel provides better peak signal-to-noise ratio (PSNR) than those with a simple bilinear interpolation method [4].

A scalable video coding (SVC) method [7], [8] can achieve the adaptation of a bit-rate with useful features such as temporal and spatial scalabilities. Scalable video coding is a promising technology for spatial, temporal, and quality scalability on a heterogeneous-network or a home-network environment. The scalable video data can be distributed anywhere at anytime, and to any device that supports needs of multimedia customers. However, SVC has various limitations such as coding performance, complex decoding, specific applications, and so on.

The H.264 [9] codec, which is a joint standard of the ITU-T video compression and the ISO/IEC MPEG-4 Part 10 Advanced Video Coding (AVC), has shown a dramatic improvement in video coding performance compared to the previous H.263 [10], MPEG-2 [11], and MPEG-4 [12] codecs. Recently, Joint Video Team of ITU-T VCEG and ISO MPEG (JVT) has been providing a scalable video model for a scalable extension of H.264/AVC [13]. It has inherited most of the building blocks of H.264 [9]. For spatial scalability, 12-tap and 6-tap linear filters are exploited in the current JSVM (joint scalable video model) for down and up-sampling with a support of arbitrary ratio spatial scalability [14], [15]. Although their computational complexity and down and up-sampling performance are suitable for utilization with scalable video coding, further improvements must be made in terms of the rate-distortion performance of scalable video coding, i.e., the performance of spatial scalability depends on the adopted down/up-sampling method. Vatis *et al.* has been made in adaptive interpolation filters for motion compensation in H.264 by using least square minimization framework [16]. Segall *et al.* also proposed an alternate way of adaptive filtering by using already determined filter coefficients [17]. Recently, adaptive estimation of filter coefficients in conjunction with down-sampling method was studied [18], [19]. Also, other efforts were made to improve the spatial scalability [20].

However, we propose new up-sampling method in spatial domain to improve spatial scalability to support dyadic scalability. DCT is used in up-sampling operation, which is crucial in SVC performance for spatial scalability. New up-sampling method use type-II DCT, which is well known DCT for image compression, because the phase change in down/up-sampling using type-II DCT is the same to the defined coordinate in current SVC recommendation [13]. Also, the combined operation of proposed up-sampling between spatial and DCT domains, which is not sequential but merged operation in generating one up-sampling matrix, makes the proposed up-sampling be employed in a simple way.

The new up-sampling method using the type-II DCT has a large degree of symmetries for efficient computation. Thus, a fast algorithm of the proposed up-sampling method is also proposed. For a performance improvement, an adaptive filtering method in the type-II DCT up-sampling is applied, which applies different weighting parameters to each DCT coefficient.

In this paper, the scalable extension of H.264 as well as the rate-distortion characteristics of the spatial scalability are

Manuscript received September 06, 2006; revised December 31, 2007. First published December 09, 2008; current version published January 30, 2009. This paper was recommended by Associate Editor G. Wen.

I. Shin was with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea. He is now with the Electronics and Telecommunications Research Institute (ETRI), Broadcasting Media Group, Daejeon 305-700, Korea (e-mail: ssi@etri.re.kr).

H. W. Park is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea (e-mail: hwpark@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TCSVT.2008.2009256

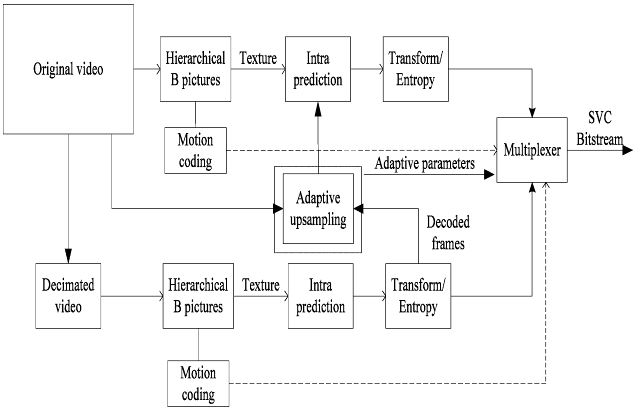


Fig. 1. Block diagram of the proposed adaptive up-sampling method for encoder.

briefly explained in Section II. The proposed up-sampling method using the type-II DCT is introduced in Section III. Section IV describes the fast algorithm of the proposed method using the symmetric properties of the DCT and IDCT kernels. The adaptive up-sampling method is also described in Section V. Experimental results are given in Section VI. Section VII concludes the paper.

II. SCALABLE EXTENSION OF H.264 AND SPATIAL SCALABILITY

The current SVC extension of the H.264 used layered spatial scalability and intrinsic quality scalability. The layered approach was used by previous MPEG-2 [11] and MPEG-4 codec [12] for supporting spatial scalability, which allowed two or three spatial layers with dyadic way [13]. The fine grain scalability (FGS) coding scheme in H.264 SVC makes rate scalability finely controlled like MPEG-4 FGS [12]. The temporal scalability can be performed by using hierarchical B pictures [13][1].

Fig. 1 shows the proposed SVC encoder structure, where two spatial layers are used for the SVC. The motion, texture, and residual data from the lower spatial layer can be reused to encode the upper spatial layer. For texture coding in the H.264 SVC, up-sampled image from lower spatial layer can be used to remove spatial redundancy. The current H.264 SVC uses up-sampling filter using convolution operation. The difference between the current spatial layer and the up-sampled signal from its lower layer is coded using a typical transform coding. The added block of the proposed method can replace the current H.264 SVC texture up-sampling filter in Fig. 1. The residual signal after motion compensation also can be used with the up-sampling operation for spatial scalable coding, and a simple bilinear up-sampling is adopted for efficient computation. The up-sampled residual signal provides prediction signal for motion compensation at the spatial enhancement layer. It should be noted that the proposed method only handle texture up-sampling. For motion data from lower layer is used with simple scaling of magnitude and block type in according to the motion vector and block type [13].

H.264 SVC support arbitrary-ratio spatial scalability, where it is called as extended spatial scalability (ESS) [27]. Also complex coordinate change problem between luminance and chrominance components was solved by allowing phase shift in down/up-sampling. In result, H.264 SVC has compliant coordinate system with H.264. H.264 SVC has texture up-sampling method with convolution, where four and two-taps poly phase filter is used for luminance and chrominance components, respectively. The filter set for up-sampling is made with integer, and approximated into 16 phases. Although accurate phase calculation is required for down/up-sampling, but approach taken in H.264 SVC is sufficient in an actual sense. Also down-sampling method adopted in JSVM [27] uses windowed sinc function, and also it is approximated into 16 phases. The down-sampling filter has characteristics for strong antialiasing, where aliasing is inevitable in down-sampling. However, further study should be done for friendly visual quality and minimization of information loss.

The rate-distortion characteristics of the spatial scalability in a video coding are determined by the base and enhancement layer bit-rates. The higher quality of the base layer leads to a poor quality of the enhancement layer. There is a trade-off relationship between the base layer image quality and the rate-distortion performance of the enhancement layer. However, the required quality of the base layer depends on the service applications. Therefore, the down/up-sampling method should be improved for the performance of the spatial scalability.

It is important to note that three factors affect the performance of the spatial scalability. These are the image quality of the lower spatial layer, the bit-rates of the spatial enhancement layer, and the characteristics of the video signal content. When the lower spatial layer has good quality, the enhancement layer quality is highly affected by the performance of the adopted up-sampling method.

III. THE PROPOSED UP-SAMPLING METHOD USING TYPE-II DCT

The type-I DCT-based up-sampling filters without phase shift [21] is proposed in a typical coordinate of down/up-sampling. However, the down/up-sampling of the current H.264 SVC suffers from a half-pel shift similar to the YUV420 down-sampling shown in Fig. 2. As the up-sampling coordinate should be matched to the down-sampling coordinate, a new up-sampling filters having a half-pel shift [22] is proposed. The proposed up-sampling method can be used in conjunction with the down-sampling filter of the current H.264 SVC, where the down-sampling filter of H.264 SVC has a half-pel shift.

The DCT family has four types, termed here types I, II, III, and IV. Each type of DCT means applied periodicity and symmetry in generating real transform with Fourier transform. For example, when symmetric extension is performed in the input signal for Fourier transform at each end of input signal, then real transform is made with property of discrete Fourier transform.

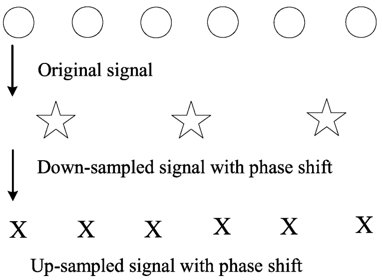


Fig. 2. SVC pixel coordinates in down/up-sampling.

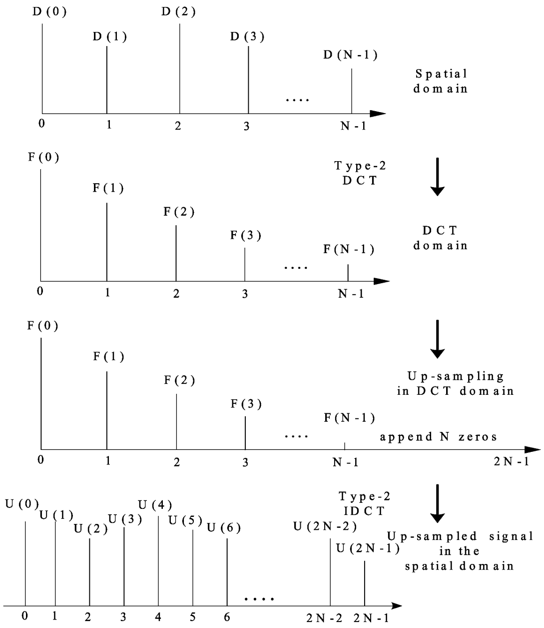


Fig. 3. Schematic two-fold up-sampling operation using DCT.

Each type of DCT has various properties and specific applications. Type-III DCT is the inverse transform of type-II DCT, which is known as IDCT. A variant of Type-IV DCT is called as Modified DCT in audio coding. The most common DCT is the type-II, which is used in the image compression due to the asymptotic equivalence of the optimal KL transform. The definition of the type-II DCT is given as follows:

$$F(k) = s(k) \sum_{n=0}^{N-1} f(n) \cos\left(\frac{\pi k(2n+1)}{2N}\right)$$

where

$$s(0) = \frac{1}{\sqrt{N}}, \quad s(k) = \frac{\sqrt{2}}{\sqrt{N}} \text{ for } 1 \le k \le N-1 \qquad (1)$$

where $f(n)$ is the input signal and N is the number of input samples. When the type-II DCT is used in down/up-sampling, the down/up-sampled coordinates have a half-pel shifted result, as shown in Fig. 2. Fig. 3 represents the up-sampling schemes using the type-II DCT in the DCT domain. The signal length of the type-II DCT is N, and N zeros are appended in the high frequency region after the type-II DCT. Following this, the type-II IDCT of the extended $2N$ samples is performed to obtain the two-fold up-sampled data shown in Fig. 3. As the cascaded operation in up-sampling using DCT and IDCT is inefficient in computational complexity, a combined operation of type-II DCT based up-sampling is proposed, where it becomes the spatial-domain up-sampling filter. It is possible to describe the two-fold up-sampling process in a matrix form as follows:

$$\begin{aligned} B^u_{2N\times 2N} &= T^t_{2N\times 2N} \times \begin{bmatrix} T_{N\times N}\cdot B_{N\times N}\cdot T^t_{N\times N} & O_{N\times N} \\ O_{N\times N} & O_{N\times N} \end{bmatrix} \\ &\quad \times T_{2N\times 2N} \\ &= \left(T^t_{N\times N}\cdot \overset{U}{T}_{N\times 2N}\right)^t \cdot B_{N\times N}\cdot \left(T^t_{N\times N}\cdot \overset{U}{T}_{N\times 2N}\right) \\ &= V^U_{2N\times N} \times B_{N\times N} \times H^U_{N\times 2N} \qquad (2) \end{aligned}$$

where $T_{N\times N}$ denotes the 1-D type-II DCT kernel for N samples, and $B_{N\times N}$ and $B^u_{2N\times 2N}$ are the original image block of $N \times N$ and the up-sampled image block of $2N \times 2N$, respectively. $\overset{U}{T}_{N\times 2N}$ represents the upper most N rows of $T_{2N\times 2N}$, and the superscript t indicates the transpose of the matrix. The inverse transform of the type-II DCT is simply the transpose of the forward type-II DCT. In (2), $O_{N\times N}$ is the $N \times N$ zero matrix. The scaling factor in the matrix calculation is omitted for simplicity. $V_{2N\times N}$ and $H_{N\times 2N}$ denote the vertical and horizontal up-sampling kernels, respectively. There is a transpose relationship between the vertical and horizontal up-sampling matrices, i.e., $V^t_{2N\times N} = H_{N\times 2N}$. Each element of the two-fold type-II DCT-based up-sampling matrix to a vertical direction can be written as follows:

$$\begin{aligned} v_{2N\times N}(n_1, n_2) = \sum_{k=0}^{N-1} p(k)\cdot \cos\left(\frac{\pi k(2n_2+1)}{2N}\right) \\ \cdot \cos\left(\frac{\pi k(2n_1+1)}{4N}\right) \end{aligned}$$

where

$$\begin{aligned} p(0) = \frac{1}{N}, \quad p(k) = \frac{2}{N} \text{ for } 1 \le k \le N-1, \\ 1 \le n_2 \le N-1, \ 0 \le n_1 \le 2N-1 \qquad (3) \end{aligned}$$

where $v_{2N\times N}(n_1, n_2)$ is the (n_1, n_2) element of the vertical up-sampling matrix, $V_{2N\times N}$. The 2-D up-sampling can be separately performed using the vertical up-sampling followed by the horizontal up-sampling. Hereinbefore we only discussed up-sampling using DCT in a dyadic way. Also, the focus of this paper is about two-fold case. The extension to an arbitrary ratio is possible with DCT, and is compliant to coordinate system of H.264 SVC. However, very complex matrix structure will be made in combining into one up-sampling kernel. The arbitrary ratio up-sampling with DCT can be made, and generalized.

IV. FAST ALGORITHM OF THE PROPOSED TYPE-II DCT UP-SAMPLING

The combined kernel of the forward and backward DCT has a large degree of symmetry [22]. The type-II DCT-based up-sampling kernel of (3) is rewritten as (4), shown at the bottom of the page. From (4), it is shown that the up-sampling kernel is symmetric, i.e., $v_{2N\times N}(n_1,n_2)=v_{2N\times N}(2N-n_1-1,N-n_2-1)$. This symmetry can be written as follows

$$V_{2N\times N}=\begin{bmatrix} v_{0,0} & \cdots & v_{0,N-1}\\ \vdots & & \vdots\\ v_{N-1,0} & \cdots & v_{N-1,N-1}\\ v_{N,0} & \cdots & v_{N,N-1}\\ \vdots & & \vdots\\ v_{2N-1,0} & \cdots & v_{2N-1,N-1}\end{bmatrix} =\begin{bmatrix} v_{0,0} & \cdots & v_{0,N-1}\\ \vdots & & \vdots\\ v_{N-1,0} & \cdots & v_{N-1,N-1}\\ v_{N-1,N-1} & \cdots & v_{N-1,0}\\ \vdots & & \vdots\\ v_{0,N-1} & \cdots & v_{0,0}\end{bmatrix} \tag{5}$$

where $v_{i,j}=v_{2N\times N}(i,j)$ for simplicity. The general symmetry in the type-II DCT is that the elements of each row are symmetrical in even rows and antisymmetric in odd rows. However, the combined kernel has a different symmetry. Therefore, a fast algorithm of the type-II DCT-based up-sampling kernel should be designed carefully. The symmetry of the proposed up-sampling kernel can be exploited to reduce the number of multiplications.

Let $V_{1,N\times N/2}$ and $V_{2,N\times N/2}$ be defined, whose (n_1,n_1)-elements are given as follows:

$$\begin{aligned} &v_{1,N\times N/2}(n_1,n_2)\\ &\quad=\frac{(v_{2N\times N}(n_1,n_2)-v_{2N\times N}(n_1,N-n_2-1))}{2}\\ &\quad=\sum_{k=0}^{N-1}p(k)\cdot\frac{(1-(-1)^k)}{2}\cdot cc(n_1,n_2,k)\\ &v_{2,N\times N/2}(n_1,n_2)\\ &\quad=\frac{(v_{2N\times N}(n_1,n_2)+v_{2N\times N}(n_1,N-n_2-1))}{2}\\ &\quad=\sum_{k=0}^{N-1}p(k)\cdot\frac{(1+(-1)^k)}{2}\cdot cc(n_1,n_2,k)\\ &\text{for } 0\le n_1<N,\quad 0\le n_2<\frac{N}{2} \end{aligned} \tag{6}$$

where $cc(n_1,n_2,k)$ is defined in (4). It is possible to decompose the up-sampling matrix using $V_{1,N\times N/2}$ and $V_{1,N\times N/2}$ as follows:

$$\begin{aligned} V_{2N\times N}\times OB_{N\times N}&=V\times\begin{bmatrix} I_{N/2} & I_{N/2}\\ -R_{N/2} & R_{N/2}\end{bmatrix}\times\begin{bmatrix} S_N\\ A_N\end{bmatrix}\times OB\\ &=\begin{bmatrix} I_{N/2} & I_{N/2}\\ -R_{N/2} & R_{N/2}\end{bmatrix}\\ &\quad\times\begin{bmatrix} V_1\\ V_2\end{bmatrix}\times\begin{bmatrix} S_N\\ A_N\end{bmatrix}\times B\\ &=\begin{bmatrix} V_1\cdot S_N+V_2\cdot A_N\\ R_N\cdot(V_2\cdot A_N-V_1\cdot S_N)\end{bmatrix}\times OB \end{aligned}$$

where

$$\begin{aligned} S_N&=[I_{N/2}\quad -R_{N/2}]\\ A_N&=[I_{N/2}\quad R_{N/2}]\\ R_N&=\begin{bmatrix} 0\,0\,0\,\ldots\,0\,0\,1\\ 0\,0\,0\,\ldots\,0\,1\,0\\ \vdots\quad\ddots\quad\vdots\\ 1\,0\,0\quad\ \ldots\end{bmatrix}. \end{aligned} \tag{7}$$

In (7), I_N and R_N are the $N\times N$ identity and antiidentity matrices, respectively, and OB is original image block to be upsampled. Since $V_{2,N\times N/2}$ is also symmetrical, i.e.,$V_{2,N\times N/2}(N-n_1-1,(N/2)-n_2-1)=V_{2,N\times N/2}(n_1,n_2)$, further decomposition of $V_{2,N\times N/2}$ is possible, where $V_{2,N\times N/2}$ is decomposed into $V_{1,N/2\times N/4}$ and $V_{2,N/2\times N/4}$ through (6). Therefore, it can be recursively decomposed until the final matrix becomes two rows and one column, where N should be a power of two. In addition, the element of the final $V_{2,N\times N/2}$ filter can be written as follows

$$V_{2,2\times 1}=\begin{pmatrix}\frac{1}{N}\\ \frac{1}{N}\end{pmatrix}=\begin{pmatrix}2^{-M}\\ 2^{-M}\end{pmatrix},\ \text{where } N=2^M. \tag{8}$$

$$\begin{aligned} v_{2N\times N}(n_1,n_2)&=\sum_{k=0}^{N-1}p(k)\cdot\cos\left(\frac{\pi k\cdot(2n_1+1)}{4N}\right)\cdot\cos\left(\frac{\pi k\cdot(2n_2+1)}{2N}\right)\\ &=\sum_{k=0}^{N-1}\frac{p(k)}{2}\cdot\left(\cos\left(\frac{\pi k\cdot(4n_2-2n_1+1)}{4N}\right)+\cos\left(\frac{\pi k\cdot(4n_2+2n_1+3)}{4N}\right)\right)\\ &=\sum_{k=0}^{N-1}p(k)\cdot cc(n_1,n_2,k)\\ \text{where } p(0)&=\frac{1}{N},\quad p(k)=\frac{2}{N}\ \text{for } 1\le k\le N-1. \end{aligned} \tag{4}$$

For computational efficiency of $V_{1,N\times N/2}$, $V_{1,N\times N/2}$ can be described as (9), shown at the bottom of the page, where $A_{N/2\times N/4}$, $B_{N/2\times N/4}$, $C_{N/2\times N/4}$, and $D_{N/2\times N/4}$ are the submatrices of $V_{1,N\times N/2}$. The element of $A_{N/2\times N/4}$ and $D_{N/2\times N/4}$ have a relationship as follows:

$$\begin{aligned} &a_{N/2\times N/4}(n_1,n_2) - d_{N/2\times N/4} \\ &\quad\times\left(\frac{N}{2}-2-2\cdot n_2, \frac{N}{4}-1-\frac{n_1}{2}\right) \\ &\quad= a_{N/2\times N/4}(n_1+1,n_2) - d_{N/2\times N/4} \\ &\quad\quad\times\left(\frac{N}{2}-2-2\cdot n_2+1, \frac{N}{4}-1-\frac{n_1}{2}\right). \end{aligned} \tag{10}$$

Additionally, elements of $B_{N/2\times N/4}$ and $C_{N/2\times N/4}$ have a relationship as follows:

$$\begin{aligned} &b_{N/2\times N/4}\left(n_1, n_2-\frac{N}{4}\right) - c_{N/2\times N/4}\left(2n_2-\frac{N}{2}, \frac{n_1}{2}\right) \\ &\quad= b_{N/2\times N/4}\left(n_1+1, n_2-\frac{N}{4}\right) \\ &\quad\quad- c_{N/2\times N/4}\left(2n_2+1-\frac{N}{2}, \frac{n_1}{2}\right). \end{aligned} \tag{11}$$

By using the relationships in (10) and (11), $V_{1,N\times N/2}$ is decomposed further for complexity reduction [22]. The horizontal up-sampling procedure can be performed in a similar way.

The proposed fast filtering method ($N = 8$) reduces the number of multiplications of the direct matrix implementation to approximately 74%, as shown in Table I. It has a comparable computational complexity to the H.264 SVC up-sampling filter. For performance improvements, we adopt an overlapping method [21]. In overlapping of two-fold up-sampling, the extended 4 points are overlapped for filtering at left and right of the current block [21]. Although the proposed fast method with overlapping is about two times slower than H.264 SVC up-sampling filter, it also shows great reduction of complexity in comparison with direct matrix multiplication. The 6×6 adaptive filter [16] shows largest complexity due to the non-separability, where it does not have symmetry in vertical and horizontal directions due to the phase shift in up-sampling.

TABLE I

NUMBER OF MULTIPLICATIONS FOR THE PROPOSED, THE JSVM UP-SAMPLING, AND THE 6 × 6 ADAPTIVE FILTER METHODS

Up-sampling method	Multiplication per pixel
Proposed without fast method (N=8)	12
Proposed with fast method (N=8)	3.1875
Proposed overlap method without fast method	21
Proposed overlap method with fast method	6.1250
H.264 SVC [13]	3
6×6 adaptive filter [16]	36

V. ADAPTIVE UP-SAMPLING METHOD USING TYPE-II DCT

The proposed up-sampling method can be adaptively applied according to the image characteristics [23]. The proposed adaptive approach also employs the type-II DCT-based up-sampling matrix, where each frequency component is weighted differently in the type-II DCT. The weighting parameters for the frequency components are adaptively determined in the proposed adaptive method, whereas these are all in unity in the fixed method described in the previous section. The weighting parameters make image to be closely to the original image by using DCT coefficient adjustment. For example, when original image has low frequency component, the weighting parameters in the DCT domain will be constant. But on the other hand, the weighting parameters depends on the quality of lower resolution image and frequency characteristics of original image.

An adaptive vertical up-sampling matrix, $V^A_{2N\times N}$, is derived as follows:

$$V^A_{2N\times N} = \left(T^U_{N\times 2N}\right)^t \cdot P_v \cdot T_{N\times N}$$

where

$$P_v = \begin{pmatrix} p_v(0)\;0 & \ldots & 0 \\ 0\;p_v(1)\;0 & \ldots & 0 \\ 0\;\;0\;p_v(2)\;0 & \ldots & 0 \\ 0 & \ldots\ldots\ldots\ldots & 0 \\ 0\;\;0 & \ldots & p_v(N-1) \end{pmatrix}. \tag{12}$$

Equation (12) shows the adaptive up-sampling kernel, where the diagonal matrix P_v represents the weighting parameters depending on the video characteristics. The horizontal and vertical up-sampling can be performed with $V^A_{2N\times N}$ and $H^A_{N\times 2N}$ matrices, respectively. The (n_1, n_2)-element of $V^A_{2N\times N}$ can be written as follows:

$$\begin{aligned} V_{1,N\times N/2} &= \begin{bmatrix} A_{N/2\times N/4} & B_{N/2\times N/4} \\ C_{N/2\times N/4} & D_{N/2\times N/4} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} a_{00} & \cdots & a_{0(N/4-1)} \\ \vdots & \ddots & \vdots \\ a_{(N/2-1)0} & \cdots & a_{(N/2-1)(N/4-1)} \end{bmatrix} & \begin{bmatrix} b_{00} & \cdots & b_{0(N/4-1)} \\ \vdots & \ddots & \vdots \\ b_{(N/2-1)0} & \cdots & b_{(N/2-1)(N/4-1)} \end{bmatrix} \\ \begin{bmatrix} c_{00} & \cdots & c_{0(N/4-1)} \\ \vdots & \ddots & \vdots \\ c_{(N/2-1)0} & \cdots & c_{(N/2-1)(N/4-1)} \end{bmatrix} & \begin{bmatrix} d_{00} & \cdots & d_{0(N/4-1)} \\ \vdots & \ddots & \vdots \\ d_{(N/2-1)0} & \cdots & d_{(N/2-1)(N/4-1)} \end{bmatrix} \end{bmatrix} \end{aligned} \tag{9}$$

$$v^{A}_{2N\times N}(n_1,n_2)=\sum_{k=0}^{N-1}p(k)\cdot p_v(k)\cdot\cos\left(\frac{\pi k(2n_2+1)}{2N}\right)\cdot\cos\left(\frac{\pi k(2n_1+1)}{4N}\right)$$

where

$$p(0)=\frac{1}{N},\quad p(k)=\frac{2}{N}\text{ for }1\le k\le N-1,\quad 1\le n_2\le N-1,\ 0\le n_1\le 2N-1 \tag{13}$$

where $p_v(k)$ denotes the adaptive parameters. When the proposed adaptive up-sampling method is applied, the up-sampling matrices of $V^{A}_{2N\times N}$ and $H^{A}_{2N\times N}$ are computed for every frame using the optimum weighting parameters of P_v and P_h, respectively.

The computational complexity of the adaptive up-sampling method is identical to that of the fixed up-sampling method, and the proposed fast algorithm in Section IV can be applied to the adaptive method. A 5-bit quantization of the weighting parameters was experimentally determined to be suitable for the transmission of the parameter to the decoder. The coding scheme of exp-golumb [9] is applied, where each weighting parameter is preprocessed by DPCM (differential pulse coded modulation). As shown in Fig. 1, the transmitted adaptive parameters are used to reconstruct the type-II DCT adaptive up-sampling kernel using (15), and the adaptive up-sampling is performed at the decoder side to provide the prediction signal on the current spatial layer.

The least-square minimization method was used for an estimation of the weighting parameters. Let the i denote the $N\times N$ block index of the down-sampled image, which has a width (W) and a height (H), then the optimum P_v and P_h that minimize error between the original up-sampled image and the computed up-sampled image are obtained, as follows:

$$\arg\mathrm{Min}_{P_v,P_h}\sum_{i=1}^{NB}\left|B^{i}_{\mathrm{orig}}-\left(T^{U}_{N\times 2N}\right)^{t}\cdot P_v\cdot T_{N\times N}\cdot B^{i}_{N\times N}\cdot T^{t}_{N\times N}\cdot P_h\cdot T^{U}_{N\times 2N}\right|^{2} \tag{14}$$

where $B^{i,s}_{\mathrm{orig}}$ and B^{i}_{orig} are the i'th block of the original up-sampled image and down-sampled image, respectively, and NB is the number of blocks in the down-sampled image. Equation (14) can be addressed as a nonlinear least square optimization. Various works were proposed in nonlinear optimization [24]–[26], including iteration methods. However, the iteration process in a global optimization can be a serious burden to a real time system. Therefore, a simple suboptimal solution using cyclic coordinate descent algorithm [24], [26] is proposed. The optimal P_h is assumed as identity matrix. Initially, the optimal P_v is obtained using the linear least square optimization method with the identity matrix P_h. At the second step, the obtained P_v is fixed to estimate P_h with the linear least square optimization. Finally, a new P_v is determined with the optimum P_h obtained in the second step. Three iterations were experimentally determined to be sufficient for the suboptimal solution. Since each iteration can be treated as a linear least-sqaure problem, the closed-form solution of optimal estimation of vertical parameters can be written as (15), shown at the bottom of the page, where $B^{i,s}_{\mathrm{orig}}$ is the stacked matrix of B^{i}_{orig}, whose size is $2N\times 2N$. In (15), $H^{t}_{2N\times N}$ is the optimally estimated horizontal up-sampling filter in previous iteration, and $\otimes$ denotes Kronecker tensor product. E^{i}_{k} and $A^{i,v}_{k}$ denote k'th column vector of each matrix, and $*$ means element-wise multiplication on matrix operation. $T^{U}_{N\times 2N}\cdot(T^{U}_{N\times 2N})^{t}$ is the diagonal matrix due to the orthogonal property of transform. Hence, diagonal matrix of $T^{U}_{N\times 2N}\cdot(T^{U}_{N\times 2N})^{t}$ can be excluded in the matrix inversion operation of (15), the computational complexity is reduced from $O(N^4)$ to $O(N^3)$ in optimal estimation. After obtaining optimal vertical parameters, the same procedure is applied for horizontal parameters in an iterative fashion.

TABLE II
EXPERIMENTAL CONDITIONS

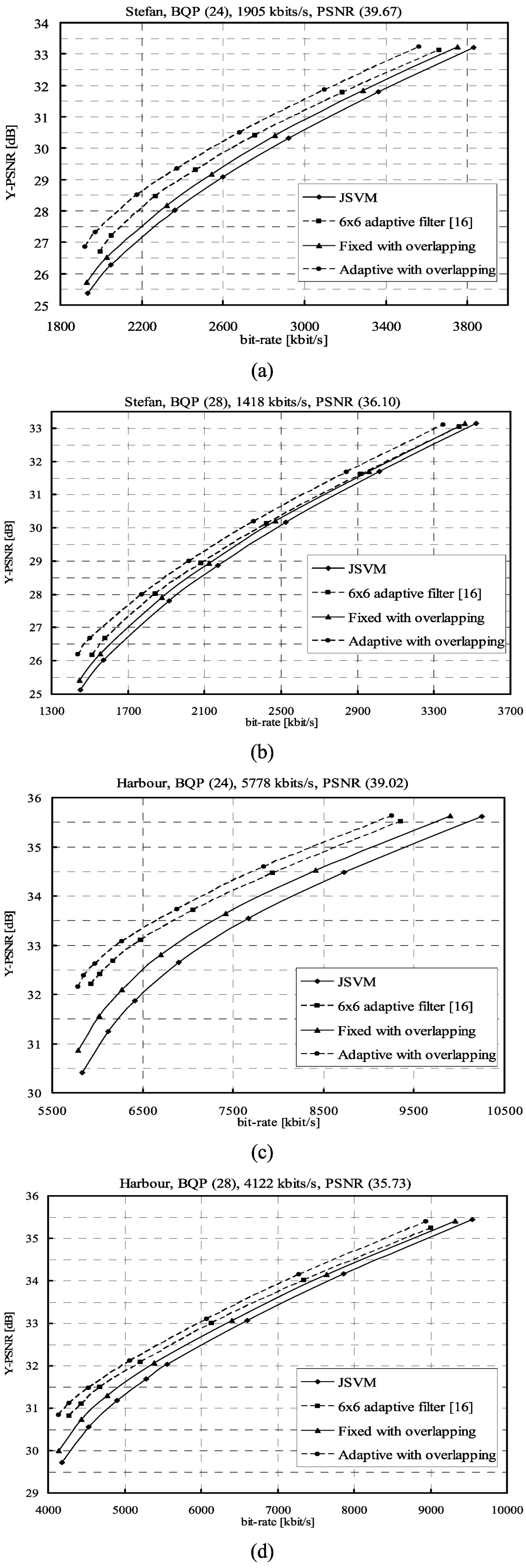
ORIGINAL SEQUENCES	STEFAN(CIF),CITY AND HARBOUR(4CIF)
DOWN-SAMPLED SEQUENCES	STEFAN(QCIF),CITY AND HARBOUR(CIF)
FRAME-RATE	30 HZ
BASE LAYER QP	24, 28
ENHANCEMENT LAYER QP	44, 38, 36, 34, 32, 30
SVC DECODING	SINGLE-LOOP DECODING MODE [13]

VI. EXPERIMENTAL RESULTS

The proposed adaptive type-II DCT-based up-sampling method was implemented in the JSVM video codec [13]. Down-sampled video was generated by a JSVM down-sampling filter. The other condition is listed in Table II.

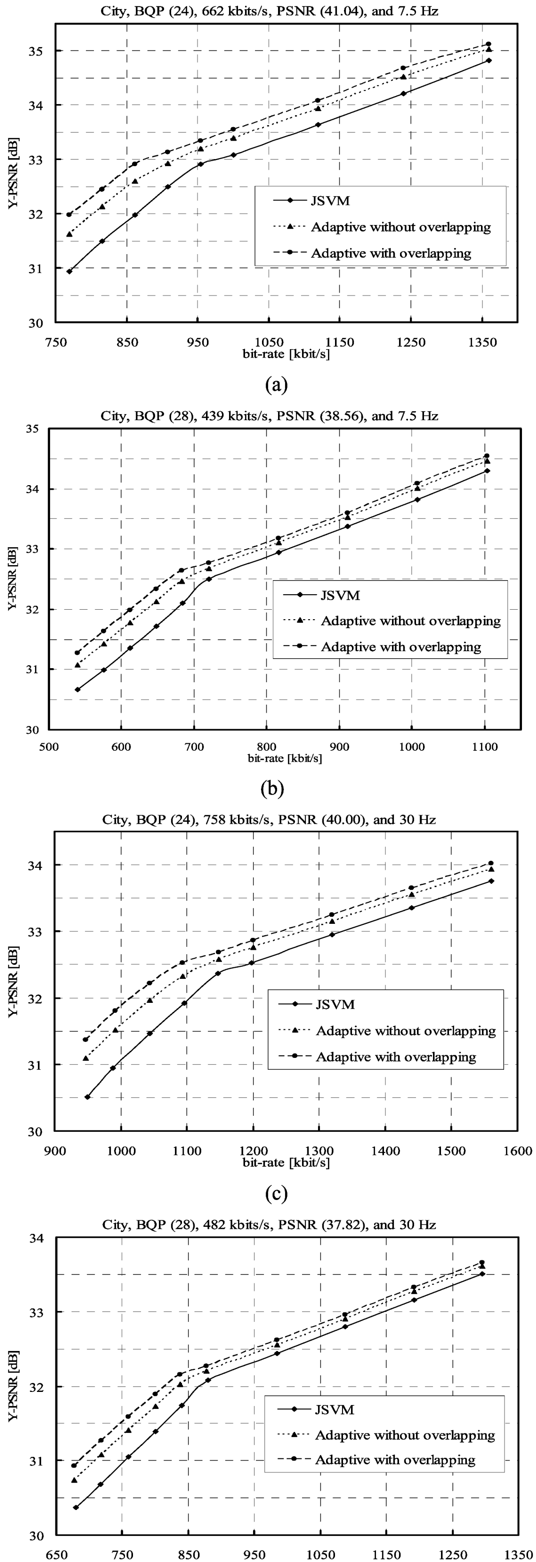
A 2-D adaptive nonseparable wiener interpolation filter [16] was implemented for performance comparison. Also current 4-tap linear filter of the JSVM was experimentally compared.

$$\begin{aligned}\mathrm{Diag}(P_v^{*})&=\left(\sum_{i=0}^{NB-1}R^{i,v,t}_{4N^2\times N}\cdot R^{i,v}_{4N^2\times N}\right)^{-1}\cdot\left(\sum_{i=0}^{NB-1}R^{i,v,t}_{4N^2\times N}\cdot B^{i,s}_{\mathrm{orig}}\right)\\&=\left(\left(\sum_{i=0}^{NB-1}\sum_{k=1}^{N}E^{i,t}_{k}\cdot E^{i}_{k}\right)*\left(T^{U}_{N\times 2N}\cdot\left(T^{U}_{N\times 2N}\right)^{t}\right)\right)^{-1}\cdot\left(\sum_{i=0}^{NB-1}R^{i,v,t}_{4N^2\times N}\cdot B^{i,s}_{\mathrm{orig}}\right)\\\text{where }E^{i}&=H^{t}_{N\times 2N}\cdot B^{i,t}_{N\times N}\cdot T^{t}_{N\times N},\quad A^{i,v}=E^{i}_{2N\times N}\otimes T^{U}_{N\times 2N},\\R^{i,v}_{4N^2\times N}&=[A^{i,v}_{0}A^{i,v}_{N+1}\cdots A^{i,v}_{N^2-1}]\end{aligned} \tag{15}$$



(a) (b) (c) (d)

Fig. 4. Rate-PSNR curves of Stefan and Harbour sequences with intra only coding. (a) Base layer QP of 24 with Stefan. (b) Base layer QP of 28 with Stefan. (c) Base layer QP of 24 with Harbour. (d) Base layer QP of 28, with Harbour.



(a) (b) (c) (d)

Fig. 5. Rate-PSNR curves of City sequence with combined scalability. (a) Extracted result of 7.5 Hz and BQP 24. (b) Extracted result of 7.5 Hz and BQP 28. (c) Extracted result of 30 Hz and BQP 24. (d) Extracted result of 30 Hz and BQP 30, (GOP is 16, and two layer of FGS were used).

All frames were coded as an intra frame to show the maximum performance of the proposed method. The rate-PSNR curves are shown in Fig. 4 for intra only coding. In Fig. 4, rate means total bit-rates of base and enhancement layers, and only luminance PSNR is considered. The fixed method with overlapping denotes the proposed type-II DCT-based up-sampling

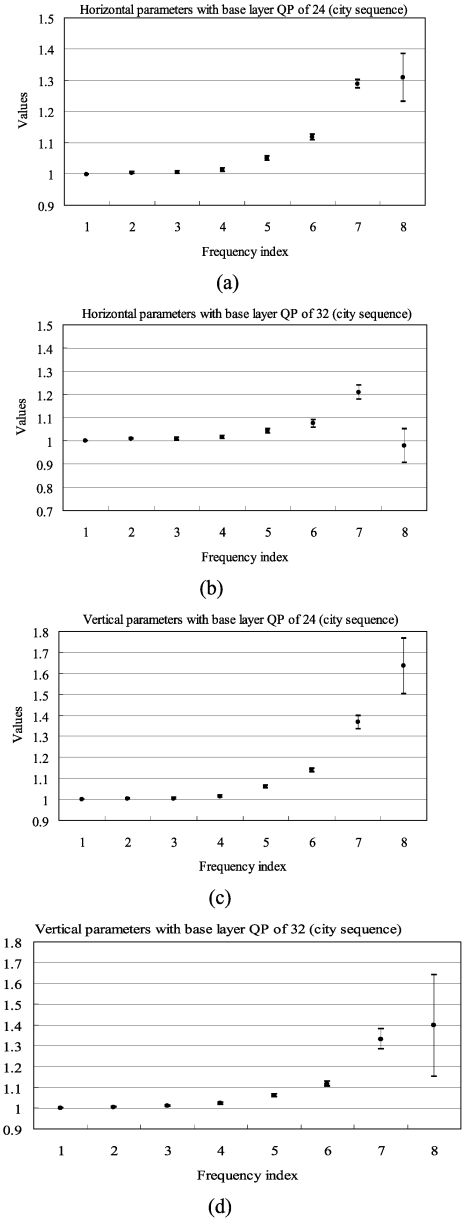


Fig. 6. Mean and standard deviation of the horizontal and vertical weighting parameters for city sequences. (a) Horizontal weighting parameters with base layer QP of 24. (b) Horizontal weighting parameters with base layer QP of 32. (c) Vertical weighting parameters with base layer QP of 24. (d) Vertical weighting parameters with base layer QP of 32.

without adaptive parameters, where overlapping method is applied as denoted in Section IV. The proposed adaptive method with overlapping shows improvements in comparison with the fixed method, where the fixed method also provides PSNR gain over JSVM method. Also, the 6×6 adaptive filter shows PSNR gain over the fixed method. However, the proposed adaptive method performs better than 6×6 adaptive filter, whose complexity is larger than adaptive method with overlapping as noted in Section IV. When the enhancement layer uses small QP or large bit-rate, any up-sampling methods do not give significant PSNR gain as discussed in Section II. When QP of the base layer increases, the PSNR gain decreases due to the incomplete down-sampled signals.

When we used intra only coding, the proposed method shows better performance than that of the 4-tap up-sampling method of JSVM. However, we consider the combined scalability with inter coding, where the combined scalability contains spatial, temporal, quality scalabilities [13]. Fig. 5 shows experimental results of inter coding, where the GOP size was 16, and experimental condition is the same as Table II.

The starting QP of FGS refinement [13] was set to 44 for providing low to high video quality. The extracted frame-rate using temporal scalability are 7.5 Hz in Fig. 5(a)–(b) and 30 Hz in Fig. 5(c)–(d), respectively.

As shown in the experimental results, the PSNR gain of the proposed method in inter coding is reduced in comparison with the result of intra only coding, because the number of texture up-sampling blocks decreases due to the small number of intra blocks in a GOP. Similar trends are shown at the high bit-rates of the enhancement layer as shown in the intra-only coding results. The proposed overlapping method improves PSNR over nonoverlapping method. Therefore, the proposed method provides much improved PSNR at the low frame rate as shown in Fig. 5.

Fig. 6 shows the mean and standard deviation of the weighting parameters for city sequences, where the weighting parameters were used adaptively for each frame. The horizontal and vertical axes show the frequency index and the optimized parameters, respectively. As shown in Fig. 6, the adaptive parameters of the horizontal and vertical up-sampling have different characteristics in each video sequence. It is also clear that the optimized weighting parameters depend on the base-layer QP. It can be understood that a different base-layer QP leads different characteristics in the frequency domain due to the quantization, even in the same image.

VII. CONCLUSION

An adaptive up-sampling method for performance improvement of spatial scalability in the H.264 SVC was proposed with a dyadic way. The up-sampling method was developed using a type-II DCT with a phase shift for correspondence with the current H.264 SVC standard. In addition, a fast algorithm was proposed for up-sampling using symmetries of the DCT kernel. By transmitting the adaptive weighting parameters of the type-II DCT-based up-sampling kernel, it led to improved results for the proposed adaptive up-sampling method in comparison with the JSVM up-sampling method. As discussed in the experimental section, proposed method provides benefits of rate-PSNR performance at the good quality of base layer and low quality of enhancement layer. When SVC coding scenario meets these circumstances, proposed method should be useful. Additionally, a further development of the arbitrary ratio up-sampling method for the support of H.264 SVC is under underway by the authors.

REFERENCES

[1] G. Keesman *et al.*, "Transcoding of MPEG bitstreams," *Signal Processing: Image Commun.*, vol. 8, pp. 481–500, 1996.

[2] B. Shen, I. K. Sethi, and B. Vasudev, "Adaptive motion vector resampling for compressed video downscaling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 9, pp. 929–936, Sep. 1999.

[3] T. Shanableh and M. Ghanbari, “Hybrid DCT/pixel domain architecture for heterogeneous video transcoding,” *Signal Processing: Image Commun.*, vol. 18, no. 8, pp. 601–620, Sep. 2003.

[4] R. Dugad and N. Ahuja, “A fast scheme for image size change in the compressed domain,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 461–474, Apr. 2001.

[5] 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, no. 9, pp. 1016–1034, Sep. 2003.

[6] C. L. Salazar and T. D. Tran, “On resizing images in the DCT domain,” in *Proc. Int. Conf. Image Processing*, Oct. 2004, pp. 2797–2800.

[7] A. T. Campbell and G. Coulson, “QoS adaptive transports: Delivering scalable multimedia to the desktop,” *IEEE Network*, vol. 11, no. 2, pp. 18–27, Mar. 1997.

[8] J. Moura, R. Jasinschi, H.-H. Shiojiri, and C. Lin, “Scalable video coding over heterogeneous networks,” *Proc. SPIE*, vol. 2602, pp. 294–306, 1996.

[9] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T and ISO/IEC JTC1, Rec. H.264—ISO/IEC 14496-10 AVC, 2003.

[10] G. Bjontegaard and K. O. Lillevo, *H.263 anchors—Technical description* MPEG95/0322, 1995.

[11] *Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at up about 1.5 Mbit/s: Video*, ISO/IEC 13818-2 (Mpeg2-Video), 1993.

[12] *Mpeg-4 Video Verification Model 8.0* ISO/IEC JTC1/SC29/WG11, MPEG97/N1796, 1997.

[13] *Joint Scalable Video Model JSVM-5,* Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-S202, Geneva, Switzerland, 2006.

[14] S. Sun, *Direct Interpolation for Upsampling in Extended Spatial Scalability*, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-P012, Poznan, Poland, 2005.

[15] S. Sun, J. Reichel, E. Francois, H. Schwarz, M. Wien, and G. J. Sullivan, U*nified Solution for Spatial Scalability* Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-R018, Bangkok, Thailand, 2006.

[16] Y. Vatis, B. Edler, D. T. Nguyen, and J. Ostermann, “Motion and aliasing compensated prediction using a two-dimensional non-separable adaptive wiener interpolation filter,” in *Proc. Int. Conf. Image Processing*, Sep. 2005, pp. 894–897.

[17] A. Segall and S. Lei, *Adaptive Upsampling for Spatially Scalable Coding,* Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-O010, Busan, Korea, 2005.

[18] A. Segall, *Study of Upsampling/Down-Sampling for Spatial Scalability* Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-Q083, Nice, France, 2005.

[19] A. Segall, *Upsampling/Down-Sampling for Spatial Scalability*, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-R070, Bangkok, Thailand, 2006.

[20] G. J. Sullivan, *Resampling Filters for SVC Upsampling*, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-R066, Bangkok, Thailand, 2006.

[21] I. H. Shin and H. W. Park, *Down/Up-Sampling Method for Spatial Scalability of SVC*, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-O024, Busan, Korea, 2005.

[22] I. H. Shin and H. W. Park, “Efficient down-up sampling using DCT kernel for MPEG-21 SVC,” in *Proc. Int. Conf. Image Processing*, Sep. 2005, pp. 640–643.

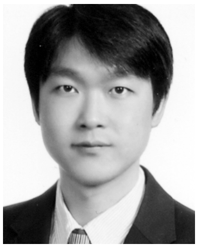
[23] I. H. Shin and H. W. Park, *Adaptive Up-Sampling Method for Spatial Scalability in SVC*, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-P045, Poznan, Poland, 2005.

[24] J. More, “The levenberg-marquardt algorithm, implementation and theory,” in *Numerical Analysis, Lecture Notes in Mathematics 630*, G. A. Watson, Ed. New York: Springer-Verlag, 1977.

[25] Z.-Q. Luo and P. Tseng, “On the linear convergence of descent methods for convex essentially smooth minimization,” *SIAM J. Contr. Optim*, vol. 30, pp. 408–425, 1992.

[26] P. Tseng, “Convergence of a block coordinate descent method for nondifferentiable minimization,” *J. Optim. Theory Appl.*, vol. 109, pp. 475–494, 2001.

[27] C. A. Segall and G. J. Sullivan, “Spatial scalability within the H.264AVC scalable video coding extension,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1121–1135, Sep. 2007.



IlHong Shin received the B.S, M.S, and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2000, 2002, and 2007, respectively.

He is currently working at the Electronics and Telecommunications Research Institute, Daejeon, Korea. His research interests include image and video coding, video communication, IPTV, and convergence of broadcasting and telecommunication system.

Hyun Wook Park (SM’99) received the B.S degree in electrical engineering from Seoul National University, Seoul, Korea in 1981 and the M.S and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Seoul, Korea in 1983 and 1988, respectively.

He has been a Professor in the Electrical Engineering Department, KAIST, Daejeon, Korea, since 1993. He was a Research Associate at the University of Washington from 1989 to 1992 and was a Senior Executive Researcher at the Samsung Electronics Co. Ltd., from 1992 to 1993. His current research interests include image computing system, image compression, medical imaging, and multimedia system. He is a member of SPIE.