

# TWO-LAYER IMAGE RESIZING FOR SCALABLE CODEC

*Ci Wang, Ping Xue and Weisi Lin\**

Nanyang Technological University Singapore, 639798 Email: {wangci, epxue}@ntu.edu.sg

\*Institute for Infocomm Research Singapore, 119613 Email: wslin@i2r.a-star.edu.sg

## ABSTRACT

Resizing compressed images is often needed for different display, storage and transmission. Based on the analysis of energy reserving capacity for different basis vectors, we propose a two-layer downsizing scheme, where the elementary layer represents the low frequency components of the image, while some high frequency details can be recovered from the enhancement layer. The new method provides a solution for low definition image scalable transmission, at the same time, provides a potential tool to rationally recover the high definition information. Experimental results show that the new scheme exhibits better picture quality than some current downsizing approaches, with a comparable computational cost.

## 1. INTRODUCTION

Nowadays, a large amount of multimedia contents is available in various sources, and the display capacities of consumer terminals also vary greatly. Changing the definition of the compressed images is a frequent task in multimedia applications. The DCT domain image resizing has gained much popularity due to its computational efficiency. As the current network structure cannot guarantee enough bandwidth, the multi-layer coding is introduced to image/video transmission. Considering the diversities of the networks and consumers, we propose a two-layer resizing scheme in DCT domain, which provides a solution for low definition images scalable transmission. Simultaneously, it has the ability for recovering the high definition information. Therefore, different from the traditional schemes, this two-layer scheme enhances not only the low definition images but also high definition images for various customs.

In the past years, Dugad and Ahuja [1] proposed a compressed image resizing method, which produced good results in terms of computational cost and image quality, and its anti-aliasing filter response was better than that of the bilinear interpolation in the spatial domain. Mukherjee

and Mitra [2] improved the method, using the subband approximation of DCT coefficients for image resizing. Wang, et. al. [3] made use of the relationship of the DCT coefficients between a block and its subblocks [4] to achieve arbitrarily image-resizing. Park, et. al. [5] carried out the image resizing operation with resizing ratio of  $L/M$  (where  $L$  and  $M$  are the upsampling and downsampling fold) by the symmetric convolution and achieved better PSNR and anti-aliasing response, at the expense of higher computational overheads. All these methods performed image downsampling by truncating DCT coefficients and upsampling by padding zero coefficients on the high frequency sides resulting in a single layer of DCT coefficients.

To keep the original information as much as possible is vital on image down-conversion. We propose a two-layer scheme, which is consisted of the elementary layer and the enhancement layer. The elementary layer is extracted from the short basis vectors using the method of Dugad and Ahuja. We further introduce an enhancement layer to include the contribution of the high frequency coefficients. The enhancement layer provides an opportunity to improve downsampled image quality when incorporated with the elementary layer. We can also reconstruct the image details from the enhancement layer by the pseudo-inverse decomposition, which improves accuracy as compared with the algorithm by Park, et. al. Compared with former multilayer scheme, which only improves the image/video quality at certain definition, the proposed scheme can provide satisfactory images to both low and high definition terminals simultaneously, and is thus especially suitable for the multicast image/video system.

The rest of this paper is organized as follows. Section 2 examines the existing resizing algorithms. A detail description of the proposed two-layer resizing strategies is given in Section 3. Experiment results with standard test images of different nature are presented in Section 4 with the comparison of anti-aliasing effects for different schemes, where we can see the anti-aliasing filter response of the proposed scheme is closer to the ideal filter than current methods. Section 5 is the conclusion.

## 2. A REVIEW OF EXISTING ALGORITHMS

---

This work is supported by I2R and NTU, Singapore, through grants LIT 2002-4 and RGM 14/02

In this section, we discuss the relevant existing resizing schemes. The energy preserving capacity is employed to evaluate down-conversion algorithms.

## 2.1. Algorithms for image halving

To perform an image halving operation on the  $8 \times 8$  block DCT space, four adjacent DCT blocks must be combined into single  $8 \times 8$  DCT block. It can be achieved in two ways. In the first approach, namely IHC (Image-Halving-Composition), each resultant  $4 \times 4$  DCT subblocks is directly [1] derived from its corresponding  $8 \times 8$  input blocks or by the subband approximation [6] for recomposing an  $8 \times 8$  DCT blocks. Let  $b_n, 1 \leq n \leq 4$  be  $8 \times 8$  image blocks in the spatial domain,  $B_n$  be the DCT blocks of  $b_n$ ,  $\hat{B}$  be the halved image in DCT domain, and  $T_n$  denote the  $n \times n$  DCT operator matrix. Four adjacent  $8 \times 8$  DCT blocks  $B_{1-4}$  are combined into an  $8 \times 8$  downsampled block  $\hat{B}$ . Dugad and Ahuja's algorithm [1] (referred as DA hereinafter) can be described by

$$\hat{B} = T_8' \begin{pmatrix} T_4' & O_4 \\ O_4 & T_4' \end{pmatrix} \begin{pmatrix} \tilde{B}_1 & \tilde{B}_2 \\ \tilde{B}_3 & \tilde{B}_4 \end{pmatrix} \begin{pmatrix} T_4 & O_4 \\ O_4 & T_4 \end{pmatrix} T_8' \quad (1)$$

where the superscript in  $T'$  denotes the transposition of  $T$ ;  $O$  is the zero matrix;  $\tilde{B}_n$  are the low frequency  $4 \times 4$  DCT subblocks, which is given by DA algorithm.

The second approach, namely ICH (Image-Composition-Halving), a  $16 \times 16$  DCT macroblock is firstly calculated from the four adjacent  $8 \times 8$  blocks, and then the  $8 \times 8$  block in the resultant image is directly derived from this  $16 \times 16$  DCT macroblock or by subband approximation. This process is expressed by

$$\hat{B} = (I_8 \quad O_8) T_{16}' \begin{pmatrix} T_8' & O_8 \\ O_8 & T_8' \end{pmatrix} \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix} \begin{pmatrix} T_8 & O_8 \\ O_8 & T_8 \end{pmatrix} T_{16}' \begin{pmatrix} I_8 \\ O_8 \end{pmatrix} \quad (2)$$

where  $I_n$  denotes the  $n \times n$  identity matrix, and the other definitions are the same as those in (1).

IHC method is operated on 8-points DCT vector, but ICH method is operated on 16-points DCT vector. There is subtle difference between these two approaches. In the first approach, truncating is done before the block recomposing, whereas in the second approach, truncating is done after block recomposing. Therefore, IHC and ICH are short basis vector (SBV) and long basis vector (LBV) truncating respectively. Both approaches have their own merits.

## 2.2. Analysis of SBV and LBV

In this subsection, we will prove both LBV and SBV cannot always get better downsampling performance than the others. MSE of the upsampling the downsampled image is used here as the criterion to evaluate the information loss.

For simplicity, we limit our discussion to 1-D DCT vectors, as the extension to 2-D is straightforward. Each of the LBVs  $x(i)$   $0 \leq i \leq 15$  is composed of two consecutive SBVs  $x_0(i), x_1(i)$ ,  $1 \leq i \leq 8$ . Let the downsampling MSE of SBVs be  $MSE_{DA}$  and the MSE of truncating LBV be  $MSE_{WA}$

The difference between  $MSE_{DA}$  and  $MSE_{WA}$  is:

$$\Delta MSE_{DA,WA} = \begin{pmatrix} x_0'(i) & x_1'(i) \end{pmatrix} (E - F) \begin{pmatrix} x_0(i) \\ x_1(i) \end{pmatrix} \quad (3)$$

where

$$E = \begin{pmatrix} T_8' A_1 T_8 & O_8 \\ O_8 & T_8' A_1 T_8 \end{pmatrix}, \quad A_1 = \begin{pmatrix} O_4 & O_4 \\ O_4 & I_4 \end{pmatrix}, \quad C_1 = \begin{pmatrix} O_8 & O_8 \\ O_8 & I_8 \end{pmatrix}$$

$F = T_{16}' C_1 T_{16}$ .  $E$  and  $F$  are related to the high frequency energy discarded during image downsampling.  $\Delta MSE_{DA,WA}$  indicates which methods is better. Some eigenvalues of the matrix  $E - F$  are positive, while the others are negative. Neither LBV truncating nor SBV truncating can always get better performance than the other. Their performances heavily depends on the contents of vectors  $x_0$  and  $x_1$ .

## 3. TWO-LAYER SCHEME FOR IMAGE RESIZING

### 3.1. Algorithms for image halving

From the orthogonal characteristics of the DCT basis vectors, the low frequency subblock is the best representation for the downsampled DCT block, but not the best one for the combined macroblock. In the proposed scheme, an original image is firstly divided into two layers: the elementary layer containing low frequency information and the enhancement layer containing high frequency information. The elementary layer is calculated by SBV truncating:

$$\hat{X}(k) = A' \begin{pmatrix} X_0(k) \\ X_1(k) \end{pmatrix} \quad A' = T_8' \begin{pmatrix} T_4' & O_4 & O_4 & O_4 \\ O_4 & O_4 & T_4' & O_4 \end{pmatrix} \quad (4)$$

We define the enhancement layer for the LBVs as

$$\Delta \hat{X}(k) = C_2 T_{16}' \Delta x(i) \quad (5)$$

$$\text{where } C_2 = (I_8 \quad O_8), \Delta x(i) = \begin{pmatrix} T_8' A_1 T_8 & O_8 \\ O_8 & T_8' A_1 T_8 \end{pmatrix} \begin{pmatrix} x_0(i) \\ x_1(i) \end{pmatrix}.$$

The halving image in DCT domain is the sum of them

$$X_h(k) = \hat{X}(k) + \Delta \hat{X}(k) \quad (6)$$

The two-layer downsampling scheme combines the merits of LBV and SBV truncating. It achieves similar low frequency response as the DA algorithm and high frequency response as the WA algorithm.  $\hat{X}(k)$  and  $\Delta \hat{X}(k)$  can be combined to achieve the minimum bitrate requirement, or transmitted separately and progressively to yield scalable transmission for better ultimate picture quality.

We can extend the 1-D analysis to the 2-D case, with

$$\hat{B}(k,l) = A \begin{pmatrix} B_1(k,l) & B_2(k,l) \\ B_3(k,l) & B_4(k,l) \end{pmatrix} A \quad \Delta\hat{B}(k,l) = C_3 \begin{pmatrix} B_1^\#(k,l) & B_2^\#(k,l) \\ B_3^\#(k,l) & B_4^\#(k,l) \end{pmatrix} C_3$$

$$B_n^\#(k,l) = \begin{pmatrix} O_8 & B_{n2} \\ B_{n3} & B_{n4} \end{pmatrix} C_3 = C_2 T_{16} \begin{pmatrix} T_8^{-1} & O_8 \\ O_8 & T_8^{-1} \end{pmatrix}$$

$$B_n(k,l) = \hat{B}(k,l) + \Delta\hat{B}(k,l) \quad (7)$$

where  $B_{n2}, B_{n3}, B_{n4}$  are the HL, LH, HH subblocks of the DCT blocks  $B_n$ , respectively.

### 3.2. Algorithms for image halving

Upsampling is the inversed process of the downsampling. An  $8 \times 8$  DCT block of the elementary layer is firstly decomposed to four  $4 \times 4$  DCT subblocks, and each of them is subsequently transformed into an  $8 \times 8$  DCT block by

$$B^\# = A\hat{B}A' \quad (8)$$

There are two approaches for recovering the high frequency part of each block from the enhancement layer signal, namely zero padding (ZP) and weighted pseudo-inverse decomposition (WPD). In the ZP approach, an  $8 \times 8$  enhancement DCT block is transformed into a  $16 \times 16$  DCT block by zero padding. It can be denoted by:

$$\Delta B^\#(k,l) = \begin{pmatrix} T_8 & O_8 \\ O_8 & T_8 \end{pmatrix} T_{16}' \begin{pmatrix} I_8 \\ O_8 \end{pmatrix} \Delta\hat{B}(k,l) \begin{pmatrix} I_8 & O_8 \\ O_8 & T_8 \end{pmatrix} T_{16} \begin{pmatrix} T_8' & O_8 \\ O_8 & T_8' \end{pmatrix}$$

This process is equal to the LBV padding. Because the enhancement layer results from the high frequency parts of the blocks  $B_n$ , it should be decomposed into the high frequency sides of  $B_n^\#$ . In the WPD approach, we adopt weighted Moore-Penrose inverse for decomposing the enhancement layer. Then, the inverse process of (7) are expressed by

$$YXY' = Z$$

$$Y = C_2 T_{16} \begin{pmatrix} T_8^{-1} & O_8 \\ O_8 & T_8^{-1} \end{pmatrix} X = \begin{pmatrix} \Delta B_1^\#(k,l) & \Delta B_2^\#(k,l) \\ \Delta B_3^\#(k,l) & \Delta B_4^\#(k,l) \end{pmatrix} Z = \begin{pmatrix} \Delta\hat{B}(k,l) & O_8 \\ O_8 & O_8 \end{pmatrix} \quad (9)$$

where  $\Delta B_n^\#(k,l)$  take the same form as  $B_n^\#(k,l)$  in (7), whose LL subband is a zero matrix. When the image is doubled, we only get the top-left block of the LBVs from its enhancement layer, i.e.  $\Delta B_n^\#(k,l)$ . We assume the coefficients in the high frequency image  $Z$ 's HL, LH and HH subbands are zeros, and the values of DCT coefficients in the HL, LH and HH subbands of the block  $\Delta B_n^\#(k,l)$  are with uniform distribution. In order to guarantee no energy of enhancement layer being mistakenly distributed into the LL bands of  $B_n^\#(k,l)$ , and  $YXY'$  in the maximum proximity to  $Z$  in low frequency bands, the weighted diagonal matrix  $M$ , whose elements corresponding to HL, LH and HH subbands of the matrix  $Z$  are zero, is introduced here for computing the approximate degree of the low frequency subbands. Weighted matrix  $N$  is designed using heavier

weighted factors on low frequency coefficients than on high frequency coefficients for measuring the MSE in  $B_n^\#(k,l)$ . The ratio of the weighted factors for the low frequency coefficients and the high frequency coefficients is 10000:1 in our simulation. We denote  $vecA$  as the vec-function, which is the vector formed by stacking the columns of the matrix  $A$  into one long vector. For WPD, the solution of (9) can be computed by the following expression

$$VecX = C_4 \cdot vec(Z) \quad C_4 = N^+ \cdot (M \cdot (Y \otimes Y) \cdot N^+)^+ \cdot M \quad (10)$$

### 4. QUALITY OF RESIZED IMAGES

The 16 samples of the input  $x(i)$  are linearly transformed with matrix  $h$  to the corresponding block of eight samples of the output  $\hat{x}(i)$ . This operation is denoted by:

$$\hat{x}(i) = h \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The  $Z$ -transform of the matrix  $h$  is given as:

$$F_k(z) = \sum_{i=0}^7 z^{-2i} H_i(z) e^{-\frac{j4\pi ki}{16}}$$

From Fig. 1, The transitional region of our  $|F_0|$  is narrower than that of the DA algorithm and IHAC algorithm in [6], and is similar to Park, et. al.'s scheme and IHCA scheme. As our scheme is the closest one to the ideal halving filter in the range  $0 \leq \omega \leq (\pi/2)$ , it has the best capability for keeping image information.

To examine the effect of the downsampling scheme, we extend the image halving to quartering utilizing the difference between SBVs (8-samples) and LBVs (32-samples). In Fig. 2, we see that the *Fencing* image obtained by our method is slightly shaper compared to the DA method and WA methods.

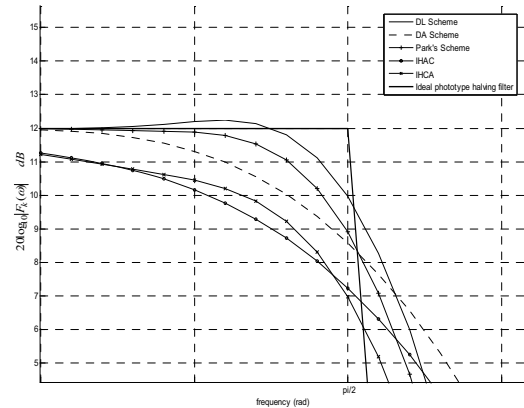


Fig 1. Comparison of frequency responses of  $F_0$  for four different image resizing

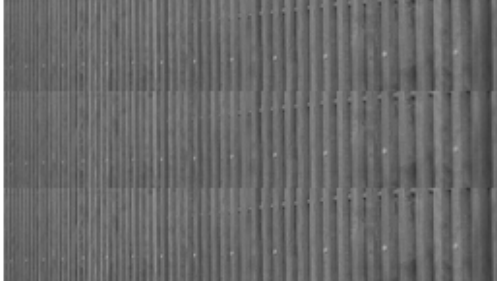


Fig 2. Downsampled Fencing images obtained by DA scheme, WA scheme and TL scheme oriented from the top to the bottom.

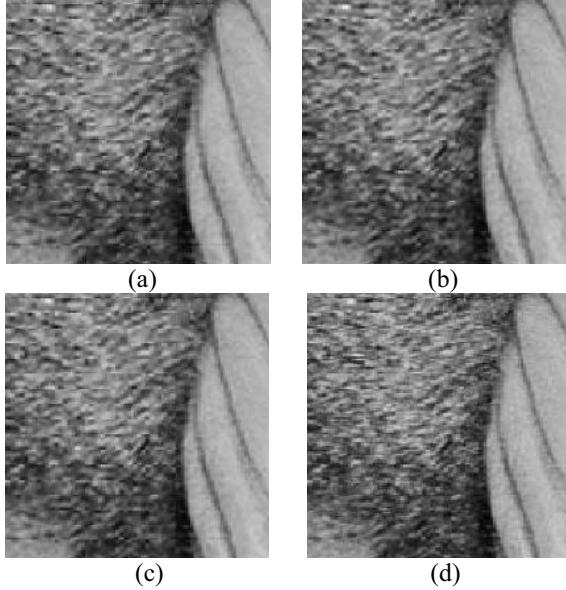


Fig 3. Parts of upsampling Baboon image after downsampled by (a) DA algorithm, (b) Park's algorithm, (c) Our algorithm by padding zero, (d) Our algorithm by pseudo-inverse compensation.

Parts of image *Baboon* are presented in Fig. 3 for visual comparison. The images obtained by DA scheme and Park scheme have some blocking artifacts along  $8 \times 8$  or  $16 \times 16$  block boundaries and the details in the blocks are blurred. On the other hand, our methods can recover some high frequency details and are with less blocking artifacts, especially in the case of TL-WPD.

Table. I illustrates the improvement of the proposed two-layer scheme with both zero padding (TL-ZP) and weighted pseudo-inverse decomposition (TL-WPD). It is observed that TL-ZP constantly outperforms the DA and the IHAC-IDAD methods about 0.3-0.4 dB; and more significant PSNR gain (about 1.5dB) has been achieved with TL-WPD over various images. Although this enhancement results from two-layer scheme rather than traditional single-layer scheme, it should be pointed out that this improvement comes from the low definition image, rather than additional data. It is most important that the quality of the low and high definition images is improved at the same time.

TABLE I. PSNR VALUES OF IMAGES THAT IS UPSAMPLING AFTER DOWNSAMPLED WITH TWO-FOLD RATIO.

Method (halving/doubling)	Lena	Boat	Peppers	Baboon
Two fold				
DA's method	34.637	31.463	32.536	24.068
Park's method	35.070	31.815	32.889	24.275
IHAC-IDAD	34.836	31.512	32.720	24.023
IHCA-IDDA	33.987	30.917	32.097	23.716
TL-ZP	35.048	31.845	32.839	24.348
TL-WPD	36.608	33.624	34.289	25.718
PSNR improvement of TL-WPD	1.538	1.809	1.4	1.443

The computational cost of the resizing algorithm is a vital factor must be considering. Referring to the computational costs of the image resizing in [6], the cost for our two-layer downsampling is similar to Wang [3], i.e. about 5 multiplications and 4.8 additions per pixel. As far as our upsampling is concerned, our TL-ZP takes 6 multiplications and 5.75 additions per pixel. Our TL-WPD has nearly two-fold higher computational cost than IHCA, and is about 10.5 multiplications and 9.7 additions per pixel.

## 5. CONCLUSION

Our scheme uses two layers to represent the downsampled image by allowing partial high frequency signal energy to be reserved as the enhancement layer. By this scheme, different users of the multicast system can get their desired halving/doubling images simultaneously, with better quality than that produced by former methods. This concept can be extended to image resizing with other ratios.

## 6. REFERENCES

- [1] 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.
- [2] J. Mukherjee, and S.K. Mitra, "Image resizing in the compressed domain using subband DCT," IEEE Trans. Circuits, Syst., Video Technol., vol. 12, pp. 620-627, Jul. 2002.
- [3] C. Wang, H.B. Yu, and M. Zheng, "A fast scheme for arbitrarily resizing of digital image in the compressed domain," IEEE Trans. Consumer Electr., vol. 49, pp. 466-471, May. 2003.
- [4] J. Jiang, and G. Feng, "The spatial relationships of DCT coefficients between a block and its sub-blocks," IEEE Trans. Signal Process., vol. 50, pp. 1160-1169, May. 2002.
- [5] H. W. Park, Y. S. Park, and S. K. Oh, "L/M-image folding in block DCT domain using symmetric convolution," IEEE Trans. Image Process., vol. 12, pp. 1016-1034, Sep. 2003.
- [6] J. Mukherjee, and S.K. Mitra, "Arbitrary resizing of images in DCT space," IEE Proc. Vis. Image Signal Process., vol. 152, pp. 155-164, Apr. 2005.
- [7] P. Lancaster and M. Timenetsky, The theory of matrices. Boca Raton, FL: Academic, 1985.