HYBRID DCT-WIENER-BASED INTERPOLATION VIA LEARNT WIENER FILTER

Kwok-Wai Hung and *Wan-Chi Siu*

Center for Signal Processing, Department of Electronic and Information Engineering
Hong Kong Polytechnic University, Hong Kong

ABSTRACT

The hybrid DCT-Wiener-based (DCT-WB) interpolation scheme provides a powerful framework to interpolate an image by utilizing the information in both spatial and DCT domain. In this paper, we investigate the bottleneck of this hybrid scheme and propose a 2D non-separable block-based Wiener filter for the hybrid scheme. The Wiener filter is learnt using training image pairs through the minimum mean squares error estimation. The proposed Wiener filter resolves the quarter-pixel shift issue and provides much better performance over the original 1D 6-tap pixel-based Wiener filter. Experimental results show that incorporating the proposed Wiener filter into the hybrid scheme improves the PSNR (0.44 dB), SSIM and subjective quality for our extensive experimental work on testing images with various contents.

Index Terms— Hybrid DCT-Wiener, up-sampling, interpolation, scalable video coding

1. INTRODUCTION

Image and video interpolation attracts much attention in the community due to the wide applications, including HDTV, video scalable coding [1-2], video zooming, surveillance, etc. For example, the 4k resolution is becoming widely available in the consumer market, such that the videos in the current DVD and HDTV formats are required to be up-sampled for displaying on the ultra high-definition TV. In recent years, many interpolation and up-sampling algorithms were proposed in the literature [3-13].

Since images and videos are compressed using block-based DCT, such as JPEG, MPEG 2 and H.264/AVC, it is straight-forward to up-sample the image in the DCT domain [7-13]. These schemes make use of the prior knowledge that the observed low-resolution (LR) image has preserved the low-frequency DCT coefficients of the original high-resolution image (HR). Due to the good energy compaction ability of DCT, the low-frequency DCT coefficients contain the most important information. Hence, the conventional DCT interpolation methods simply pad zeros as the high-frequency DCT coefficients [7-9] and provide better results than the conventional bilinear interpolation method for down-sampling and up-sampling [9].

Recently, more DCT-based interpolation methods were proposed in the literature [10-13]. They aim at improving the performance of the zero padding approaches [7-9] that introduce ringing and blocking artifacts through several means. Cho and Lee proposed the most direct way to estimate the high-frequency DCT coefficients from the low-frequency DCT coefficients through a training process [10]. However, the experimental results show that the training images need to be relevant to the testing image. An alternative idea was proposed by Shin and Park [11] to compute the adaptive weights of the low-frequency DCT coefficients, such that the errors between up-sampled image (by zero padding) and the original image is minimized. This method was proposed for the video coding, such that the original image is required to compute the weights, which are encoded. Another study by Park and Jeong [12] shows that vertical and horizontal edges can be represented by block sums, such that ringing and blocking artifacts can be largely reduced.

In this paper, we are interested in the hybrid DCT-Wiener-based interpolation method [13]. This method combines the advantages of the spatial interpolation and DCT interpolation. Since the low-frequency information of DCT is given in the low-resolution image, the performance of this hybrid scheme depends on the performance of the spatial interpolation method. We investigated the performance of the 1D 6-tap pixel-based Wiener filter used in the scheme [13] and found that quarter-pixel shift exists after up-sampling in the spatial domain. Hence, we propose a new 2D non-separable block-based Wiener filter for this scheme. Specifically, the Wiener filter is computed using some training image pairs through the minimum mean squares error estimation. Hence, the quarter-pixel shift issue in the hybrid scheme can be resolved by using the proposed Wiener filter. Experimental results show significant improvements in PSNR (0.44 dB), SSIM and subjective quality after incorporating the proposed Wiener filter into the hybrid scheme.

The rest of the organization of this paper is as follows. Section 2 briefly explain the dyadic DCT down-sampling process [7-9]. Section 3 introduces the hybrid DCT-Wiener-based interpolation framework [13]. Section 4 gives the formulation of the proposed Wiener filter. Section 5 shows some experimental results and section 6 concludes the paper.

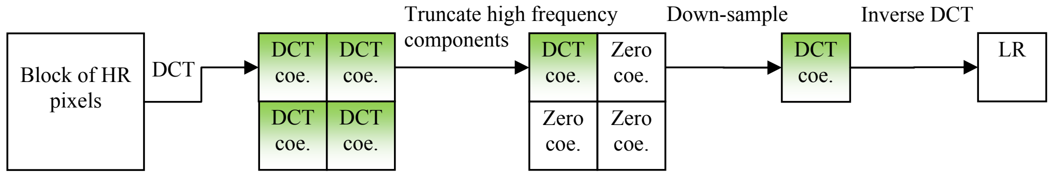


Figure 1. The graphical illustration of the dyadic down-sampling process in the DCT domain [7-9]

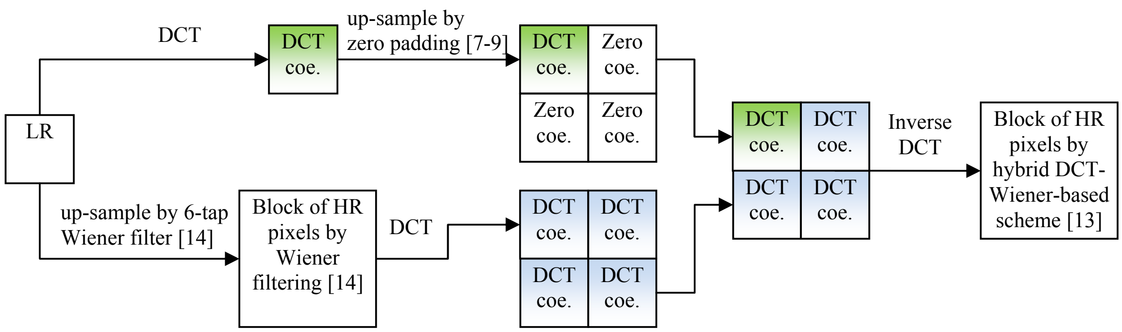


Figure 2. The graphical illustration of the hybrid DCT-Wiener-based interpolation scheme [13]

2. DCT-BASED DOWN-SAMPLING [7-9]

In order to understand better the hybrid DCT-Wiener-based interpolation scheme, we use figure 1 to illustrate the dyadic DCT down-sampling process [7-9]. Consider a block of HR pixels (e.g. 8×8). Its low-frequency DCT coefficients are kept after down-sampling. However, as explained in the introduction section, the block of pixels are always available in the DCT domain (due to JPEG, H.264/AVC formats, etc). Transforming HR pixels into the DCT domain in figure 1 is always not required in real applications.

3. HYBRID DCT-WIENER-BASED SCHEME [13]

Let us use figure 2 to illustrate the hybrid DCT-Wiener-based interpolation scheme. Assume that the low-resolution block is available in the spatial domain, then its DCT components are used to combine with the DCT components of its up-sampled version in the spatial domain. The combined frequency block is then inverse DCT transformed to give the final HR block in the spatial domain.

3.1. Drawbacks of the 1D 6-tap Wiener filter [14]

In the original scheme [13], it uses a 1D 6-tap Wiener interpolation filter for the spatial up-sampling, which is adopted in H.264/AVC as a half-pixel interpolation filter [14]. Let us consider the spatial coordinates of the LR and HR blocks after DCT-based down-sampling, as shown in figure 3. If we use the 1D 6-tap Wiener filter to interpolate the HR pixels using the LR pixels, the interpolated HR pixels have a quarter-pixel shift with the original HR pixels. Hence, the performance of the 1D 6-tap Wiener filter is 3 dB worse than the DCT up-sampling method [9], as shown in the experimental results of the paper [13] and this paper.

The best way to improve the performance of the hybrid scheme, which has a high potential to be the best DCT up-sampling scheme, is to use learning to formulate a new fixed coefficient Wiener filter specifically for the DCT down-sampled LR images to perform up-sampling. In the next section, we will introduce the formulation of the training and estimation framework of the 2D non-separable and block-based Wiener filter, which provides a much better performance than the original 1D 6-tap, quarter-pixel shifted and pixel-based Wiener filter.

4. PROPOSED LEARNT WIENER FILTER

4.1. The training framework

In the spatial domain, let us denote $\mathbf{x}$ ($n\times m$) as the training low-resolution image and $\mathbf{y}$ ($2n\times 2m$) as the training high-resolution image, where n and m are the dimensions of the image. We partition the low-resolution image into $k\times k$ blocks, and locate the corresponding $k\times k$ blocks in the high-resolution image, as illustrated in figure 4 when k=6. Then, we vectorize the blocks and group the vectors into matrices, denote $\mathbf{X}$ ($k^2\times nm/k^2$) as the low-resolution block and $\mathbf{Y}$ ($k^2\times nm/k^2$) as the high-resolution block.

The classic Wiener filter that minimizes the linear mean squares error is given by [15-20]

$$\mathbf{H} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} \tag{1}$$

where $\mathbf{H}$ is the Wiener filter coefficient matrix, which has a size of ($k^2\times k^2$). Hence, the proposed Wiener filter is a block-based filter, which use the input $k\times k$ LR pixels to estimate the $k\times k$ HR pixels, as illustrated in figure 4. In this paper, we use the Lena image (512×512) as the training image to learn the fixed coefficients of the Wiener filter. The value of k is empirically optimized to be 6.

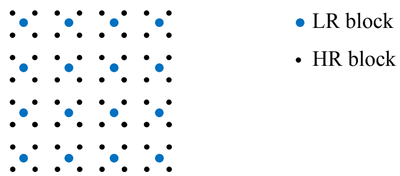


Figure 3. Spatial coordinates of the 4×4 LR block (obtained by DCT-based down-sampling [7-9]) and the 8×8 HR block.

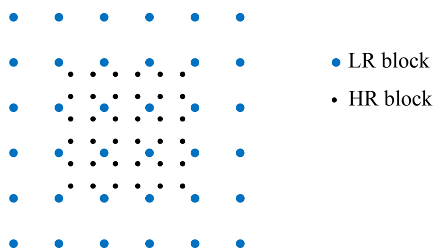


Figure 4. Spatial coordinates of the $k\times k$ LR block and the corresponding $k\times k$ HR block of the proposed Wiener filter when k=6.

4.2. The estimation framework

Given a low-resolution image **x'** ($n\times m$), where n and m are the dimensions of the image. We partition the low-resolution image into $k\times k$ overlapping blocks since our Wiener filter is a block-based filter, which is essentially different from the pixel-based Wiener filters in the literature [14-20]. Then, we vectorize the LR blocks and group the vectors into matrices, denoted as the low-resolution block **X'** ($k^2\times nm$).

Using the learnt Wiener filter in (1), we can obtain an up-sampled high-resolution block **Y'** ($k^2\times nm$) as follows

$$\mathbf{Y'} = \mathbf{HX'} \tag{2}$$

where **Y'** has a larger size ($k^2\times nm$) than the size ($k^2\times 4nm/k^2$) of the HR block of the original HR image due to overlapping when k>2. Hence, we can find the overlapped pixels in the HR block **Y'** and take the average value of overlapped pixels as the final value of the up-sampled image. The number of overlapping is given by k^2, which depends on the value of k. To better understand this, let us refer to figure 4. When k=6, the HR block overlaps for 9 times when we shift the LR block in terms of integer values. Hence, we have 9 estimated values of the HR block due to 9 overlapping using the proposed block-based Wiener filter.

4.3. Difference to adaptive Wiener filters [15-20]

One may discover that a lot of adaptive Wiener filter were proposed in the literature for the video coding [15-20]. These filters are optimal in the sense of minimizing the mean squares errors, apparently the same as the proposed Wiener filter. However, these filters are adaptive filters, which require the original HR image to compute the filter coefficients or filter parameters, such that they are applicable in the video coding only. In this paper, our Wiener filter is not an adaptive Wiener filter that has fixed filter coefficients, simply because the applications of DCT up-sampling is not limited to the video coding, when the original HR image is unavailable.

Another major difference with the available Wiener filters [14-20] is that the proposed Wiener filter is a block-based filter, which has a large block size of $k\times k$, such that the estimated blocks overlap. In our experiments, we find that this novel structure is very beneficial to the performance of the class of fixed coefficients Wiener filter. In the next section, we will show the very highly competitive performance of the proposed Wiener filter. In the future, we will further analyze the performance of this block-based structure.

5. EXPERIMENTAL RESULTS

Our experimental work was done on a Intel 3G machine. Nine testing images (512×512) of various contents as shown in figure 5 were used to verify the performance of proposed learning-based Wiener filter and the hybrid scheme [13] incorporating the proposed Wiener filter. The testing images were down-sampled by a factor of two using the DCT-based methods [7-9], as illustrated in figure 1. Block size of the HR block during the down- and up-sampling is 8×8.

Table 1 shows the PSNR (dB) and SSIM [21] comparison of various methods [7, 13, 14] for up-sampling two times. The results show that the proposed Wiener filter outperforms the DCT up-sampling method [7], the 1D 6-tap Wiener filter [14] and the original hybrid scheme [13] in terms of PSNR and SSIM. Note that we cannot compare it with the adaptive Wiener filters [15-20] directly since they require the original HR image to compute the filter coefficients. Specifically, the 3 dB advantage against the 1D 6-tap Wiener filter partly comes from resolving the quarter-pixel shift and is partly due to the new block-based structure of the Wiener filter.

Let us look at the performance of the proposed hybrid scheme (hybrid scheme [13] using the proposed Wiener filter). It provides the highest average PSNR and SSIM, and it has around 0.44 dB gain against the original hybrid scheme [13]. This result confirms the performance of the proposed Wiener filter and the benefit of the hybrid interpolation scheme, which combines the frequency information of the spatially up-sampled image and the LR image. In other words, the hybrid scheme utilizes the prior knowledge that the LR image is DCT down-sampled, such that the low-frequency components are always preserved.

Figure 6 shows the subjective comparison of various methods. The figure shows that the proposed hybrid scheme provides the sharpest picture, which is closest to the original HR image. The proposed Wiener filter provides a slightly blurry picture and the 1D 6-tap Wiener filter produce a blurry picture. The original hybrid scheme [13] produces some overshoots (pointed by blue arrows) and ringing effects around the edges (pointed by red arrows), the DCT up-sampling method has some obvious blocking artifacts (pointed by green arrows). The proposed hybrid scheme produces the most pleasant and natural images overall.

Figure 5. The testing images used in the experimental section.

Table 1: The PSNR (dB) and SSIM [21] of the testing images.

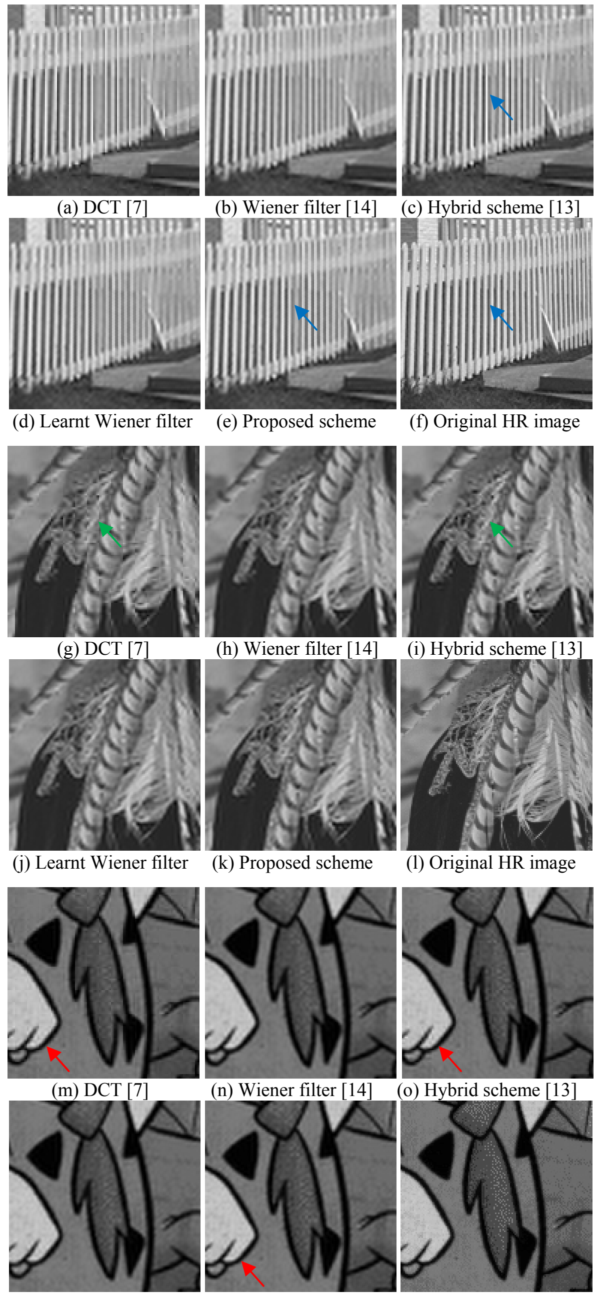
Images	DCT [7]	6-tap Wiener filter [14]	Hybrid scheme [13]	Proposed Wiener filter	Hybrid scheme using proposed Wiener filter
Baboon	24.799 0.7676	23.451 0.6904	24.755 0.7682	24.826 0.7675	**24.927** **0.7782**
Boat	30.415 0.8694	27.730 0.8207	30.429 0.8682	30.727 0.8713	**30.893** **0.8752**
Bridge	26.981 0.8259	24.953 0.7526	26.921 0.8255	27.150 0.8305	**27.283** **0.8383**
Flin	27.458 0.8753	23.124 0.8180	27.924 0.8777	28.374 0.8831	**28.734** **0.8877**
Lighthouse	26.831 0.8727	24.842 0.8302	26.877 0.8733	27.216 0.8756	**27.384** **0.8807**
Man	31.539 0.9002	28.836 0.8538	31.650 0.9014	31.863 0.9047	**32.037** **0.9084**
MRI	35.242 0.9870	31.546 0.9519	35.470 0.9851	35.480 0.9846	**35.675** **0.9866**
Pepper	32.779 0.8991	29.531 0.8713	32.339 0.8961	32.931 0.8995	**33.073** **0.9014**
Splash	39.780 0.9719	35.713 0.9602	40.023 0.9715	40.197 0.9738	**40.364** **0.9744**
Average	30.647 0.8855	27.747 0.8388	30.710 0.8852	30.974 0.8878	**31.152** **0.8923**

6. CONCLUSION

In this paper, we present a new hybrid interpolation scheme using a proposed learning-based Wiener filter for image up-sampling. The proposed Wiener filter is obtained from learning some training image pairs and has a novel block-based structure. Experimental results show that the proposed Wiener filter is highly competitive with the state-of-the-art up-sampling methods in the DCT domain. After incorporating the proposed Wiener filter into the powerful hybrid interpolation scheme, the scheme improves PSNR (0.44 dB), SSIM and subjective quality.

In the future, we can further analyze the performance of the proposed block-based Wiener filter and apply the improved hybrid scheme into the scalable video coding, since it is anticipated that the improved hybrid scheme will improve the coding efficiency due to its highly competitive performance in the general applications.

Acknowledgement: This work is supported by the Center for Signal Processing, the Hong Kong Polytechnic University and the Research Grant Council of the Hong Kong SAR Government (PolyU 5278/08E).



(a) DCT [7] (b) Wiener filter [14] (c) Hybrid scheme [13]

(d) Learnt Wiener filter (e) Proposed scheme (f) Original HR image

(g) DCT [7] (h) Wiener filter [14] (i) Hybrid scheme [13]

(j) Learnt Wiener filter (k) Proposed scheme (l) Original HR image

(m) DCT [7] (n) Wiener filter [14] (o) Hybrid scheme [13]

(p) Learnt Wiener filter (q) Proposed scheme (r) Original HR image

Figure 6. Subjective comparison of various approaches.

7. REFERENCES

[1] X. Wu, X. Zhang and X. Wang, "Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion", *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 552-561, Mar. 2009.

[2] Z. Shi, X. Sun and F. Wu, "Spatially Scalable Video Coding For HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol.22, no.12, pp.1813-1826, Dec. 2012.

[3] Wing-Shan Tam, Chi-Wah Kok and Wan-Chi Siu, 'Modified edge directed interpolation for images', *J. Electron. Imaging*, 2010, 19, (1), p. 013011

[4] He He and Wan-Chi Siu, 'Single image super-resolution using Gaussian process regression'. *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR 2011)*, 20–25 June 2011, pp. 449–456

[5] Kwok-Wai Hung and Wan-Chi Siu, "Robust soft-decision interpolation using weighted least squares," *IEEE Trans. Image Process.*, vol.21, no.3, pp.1061-1069, March 2012.

[6] Kwok-Wai Hung and Wan-Chi Siu, "Fast image interpolation using bilateral filter", *IET Image Processing*, vol. 6, no. 7, pp. 877-890, October 2012.

[7] 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.

[8] J. Mukherjee and S.K. Mitra, "Image resizing in the compressed domain using subband DCT," *IEEE Trans. Circuits Syst. Video Technol.* , vol.12, no.7, pp.620-627, Jul 2002

[9] 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 Process.*, vol. 12, no. 9, pp. 1016–1034, Sep. 2003.

[10] Min-Kyoung Cho and Byung-Uk Lee, "Discrete cosine transform domain image resizing using correlation of discrete cosine transform coefficients", *Journal of Electronic Imaging* 15(3), 033009 (Jul–Sep 2006)

[11] I. Shin and H. W. Park, "Adaptive up-sampling method using DCT for spatial scalability of scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, Feb. 2009.

[12] Sang-Jun Park and Jechang Jeong, "Hybrid image upsampling method in the discrete cosine transform domain," *IEEE Trans. Consumer Electronics*, vol.56, no.4, pp.2615-2622, November 2010

[13] Zhenyu Wu, Hongyang Yu and Chang Wen Chen, "A New Hybrid DCT-Wiener-Based Interpolation Scheme for Video Intra Frame Up-Sampling," *IEEE Signal Processing Letters*, vol.17, no.10, pp.827-830, Oct. 2010

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

[15] Wenhao Zhang, Aidong Men and Pinhua Chen, "Adaptive inter-layer intra prediction in scalable video coding," *Proceeding of IEEE International Symposium on Circuits and Systems*, pp.876-879, 24-27 May 2009

[16] Jie Dong and King Ngi Ngan, "Parametric interpolation filter for HD video coding," *IEEE Trans. Circuits Syst. Video Technol.*, Vol.20, No.12, pp.1892-1897, Dec. 2010.

[17] Y. Vatis and J. Ostermann, "Adaptive interpolation filter for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 179–192, Feb. 2009.

[18] D. Rusanovskyy, K. Ugur, A. Hallapuro, J. Lainema and M. Gabbouj," Video coding with low-complexity directional adaptive interpolation filters," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp 1239–1243, Aug. 2009.

[19] D. Rusanovskyy, K. Ugur and M. Gabbouj, "Adaptive interpolation with flexible filter structures for video coding," in *Proceeding of IEEE International Conference on Image Processing*, pp. 1025–1028, Nov. 2009.

[20] M. Karczewicz, Y. Ye, P. Chen and G. Motta, Experimental Results of Interpolation Filters on High-Definition Sequences, ITU-T Q.6/SG16 (VCEG) document VCEG-AJ30, Oct. 2008.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error measurement to structural similarity," *IEEE Trans. Image Process.,* vol. 3, no. 4, pp. 600-612, Apr. 2004.