

Design of Optimal Decimation and Interpolation Filters for Low Bit-Rate Image Coding

Wu-Sheng Lu and Ana-Maria Sevcenco
Department of Electrical and Computer Engineering
University of Victoria
Victoria, BC, Canada V8W 3P6
wslu@ece.uvic.ca, sevcenco@engr.uvic.ca

Abstract— The DCT-based JPEG standard is certainly one of the most successful applications of transform coding methods for still digital images. A commonly recognized disadvantage of the DCT-based transform coding is its blocking artifacts which become increasingly visible when the bit-rate gets lower. In the past, considerable research endeavors had been made to deal with this problem. Motivated by a recent work of Tsai, Elad, Milanfar, and Golub (TEMG), in this paper we investigate several techniques for the design of optimal decimation and interpolation filters that can be utilized in a TEMG type system framework for further performance improvement, where the DCT-based JPEG standard is used in conjunction with a decimation filter and an interpolation filter. Simulation results are presented to demonstrate the effectiveness of the proposed techniques.

I. INTRODUCTION

Despite the emerging wavelet-based standard JPEG-2000 for still images [1], the discrete cosine transform (DCT) based JPEG standard [2]-[4] remains to be one of the most successful applications of transform coding methods for still digital images.

A commonly realized disadvantage of the DCT-based transform coding is its blocking artifacts which become increasingly visible when the bit-rate gets lower. In the past, considerable research endeavors have been made to deal with this problem. The work of particular interest and relevant to the methods described below are the algorithms recently proposed in [5][6] that are JPEG-based and incorporate an anti-aliasing filtering and down-sampling pre-processing step and an interpolative up-sampling post-processing step, which have demonstrated considerable performance improvement in terms of coding gain and reduced blocking effects. Studies of image coding problems employing similar frameworks can also be found in [7]-[10].

Motivated by a recent work of Tsai, Elad, Milanfar, and Golub (TEMG) [6], in this paper we investigate several techniques for the design of optimal decimation and interpolation filters that can be utilized in a TEMG type system framework for further performance improvement, where the DCT-based JPEG standard is used in conjunction

with a decimation filter and an interpolation filter. Simulation results are presented to demonstrate the effectiveness of the proposed techniques.

II. THE METHOD OF TSAIG, ELAD, MILANFAR, AND GOLUB

As illustrated in Fig. 1, the method of TEMG [6] starts by applying anti-aliasing filtering to an input digital image X of size $m \times n$. Typically this is a separable FIR two-dimensional (2-D) filter whose transfer function can be expressed as $H(z_1, z_2) = h(z_1) \cdot h(z_2)$ where $h(z)$ is a one-dimensional (1-D) FIR lowpass transfer function with normalized cutoff frequency f_c . The output of filter H is down-sampled by k in both dimensions and the down-scaled image Y is encoded by DCT plus certain quantization and coding techniques before transmission. The received 2-D signal is decoded by corresponding decoding techniques and the resulting image \hat{Y} is then up-sampled by k and filtered by an interpolation 2-D filter G to produce the reconstructed image \hat{X} .

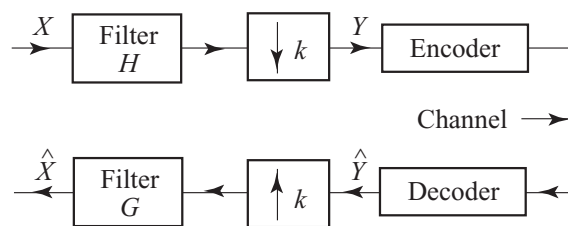


Figure 1. A TEMG type system set-up.

In [6], the scaling factor is set to $k = 2$ and the up-sampling and interpolative filtering are treated as a unified process which is performed using one filter G for interpolation. In a scenario such as this, filter G may be explicitly described by four subfilters $G^{(p,q)}$ with p and q assuming the values of 0 or 1, where each subfilter is applied to the decoded image \hat{Y} to produce a subimage of size $(m/2) \times (n/2)$, and the four filtered subimages are used to construct image \hat{X} as its even row/even column, even row/odd column, odd row/even column, and odd row/odd column parts,

respectively. Each of the four nonseparable 2-D FIR filters is obtained by minimizing the Euclidean norm of an error vector that measures the difference between an subimage $x^{(p,q)}$ obtained by down-scaling by 2 the input image X and an subimage obtained by filtering Y using subfilter $G^{(p,q)}$. Here $x^{(p,q)}$ denotes a column vector formed by taking each row of the subimage, transposing it to a column, stacking it with the next column vector, and so on. In analytic terms, let the size of each subfilter $G^{(p,q)}$ be $l \times l$, and let Φ be a matrix of size $s \times l^2$ with $s = mn/4$, where each row is obtained by associating a mask of size $l \times l$ with each component of image \hat{Y} and stacking the elements of \hat{Y} within the mask as a row vector. If we denote the vector version of the subfilter $G^{(p,q)}$ by $g^{(p,q)}$, then $g^{(p,q)}$ is obtained by minimizing the norm

$$\|\Phi \cdot g^{(p,q)} - x^{(p,q)}\|_2$$

which leads to

$$g^{(p,q)} = \Phi^+ \cdot x^{(p,q)}$$

where Φ^+ denotes the pseudo-inverse of matrix Φ . Once $g^{(p,q)}$ for $p, q \in \{0, 1\}$ are determined, they are used to produce the even row/even column, even row/odd column, odd row/even column, and odd row/odd column parts of the output image \hat{X} as

$$\hat{x}^{(p,q)} = \Phi \cdot g^{(p,q)}$$

Concerning the anti-aliasing filter H , the method in [6] proposes to use $H(z_1, z_2) = h(z_1) \cdot h(z_2)$ where $h(z)$ is a Hamming window based linear phase FIR filter whose cutoff frequency f_c is optimized so as to maximize the peak signal-to-noise ratio (PSNR).

III. OPTIMIZATION OF INTERPOLATION FILTER G

As described in Section II, the interpolation filter G in [6] is already optimal in the least squares sense, and the solution depends critically on matrix Φ . In this section, we propose two approaches for improving the design of filter G via even symmetric and odd symmetric extensions of image \hat{Y} that modify the way matrix Φ is constructed so as to improve the PSNR for a given bit rate.

A. An Even Symmetric Extension of \hat{Y}

Let $m_1 = m/2$, $n_1 = n/2$ define the size of matrix \hat{Y} , and let the size of the impulse response of filter G be $l \times l$. For simplicity, we assume that l is an odd integer and let $l_1 = (l-1)/2$. We propose a symmetric extension for image \hat{Y} as an image of size $(m_1 + 2l_1) \times (n_1 + 2l_1)$ which is obtained by extending image \hat{Y} by adding l_1 lines along each side of the image by flipping l lines of \hat{Y} nearest the image's boundary. The extension is then completed by filling in the four $l_1 \times l_1$ corners by symmetric flipping the corresponding four corners of image \hat{Y} . The matrix Φ in this case is constructed in a way similar to that described in Section II: to start, we

place the centre of the mark (of size $l \times l$) at the upper-left corner of \hat{Y} , then scan the entire \hat{Y} in the same way as in [6]. Because of the image extension, the resulting matrix Φ differs from that in [6] and in turn leads to a modified set of subfilters $G^{(p,q)}$. As an example, Fig. 2 shows the PSNR improvement in the proposed even-symmetric extension relative to that by zero-padding. The image used in the simulation was *barbara* of size 512×512 and the bit rate is in the range of 0.07 bpp to 0.19 bpp. It was found that the improvement in PSNR varies from 0.1246 dB to 0.1862 dB. The even-symmetric extension was also applied to several digital images of same size and similar amount of PSNR improvement was observed for all the images tested.

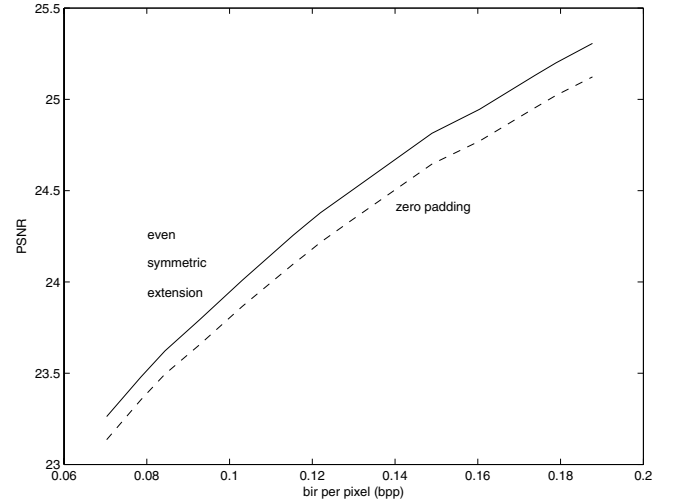


Figure 2. PSNR improvement by even-symmetric extension for image *barbara*.

B. An Odd Symmetric Extension of \hat{Y}

An odd symmetric extension may be described as a symmetric extension as in Section II.A followed by a step in that the pixel value at a given position in an extension strip is modified by adding a value to it. This value is equal to twice of the difference between the pixel value at the nearest boundary point and the pixel value at the position in question. In doing so, the image \hat{Y} is extended in smoother way in the sense that it is kind of “differentiable extension”. Fig. 3 shows the PSNR improvement in the proposed odd-symmetric extension relative to that of zero-padding. The image used in the simulation was also *barbara* of size 512×512 within the bpp range of 0.07 bpp to 0.19 bpp. It was found that the improvement in PSNR varies from 0.1236 dB to 0.1886 dB, which is quite similar to that obtained by symmetric extension. The odd-symmetric extension was also applied to several digital images of same size and similar amount of PSNR improvement was observed for all the images tested.

Based on the above, we conclude that either even or odd symmetric extension of image \hat{Y} is a worthwhile to implement because of its simplicity and good improvement in PSNR. In the next section, we turn our attention to the

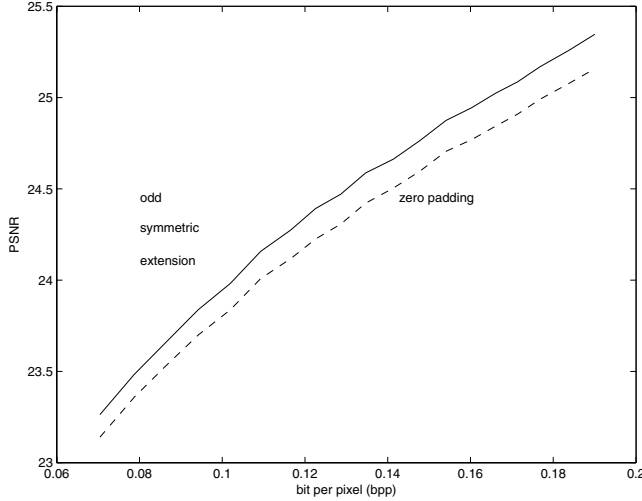


Figure 3. PSNR improvement by odd-symmetric extension for image *barbara*.

optimization of the decimation filter H .

IV. OPTIMIZATION OF DECIMATION FILTER H

The problem of optimal design of the decimation filter H is to find the coefficients of two 1-D FIR filters $h_1(z_1)$ and $h_2(z_2)$ such that the use of filter $H(z_1, z_2) = h_1(z_1) \cdot h_2(z_2)$ leads to maximum PSNR subject to a given bpp. Note that, unlike in [6], here we consider a separable 2-D filter that is composed of two *distinct* 1-D FIR filters, each with possibly *nonlinear* phase response. In other words, if we assume the two 1-D filters have the same length l_h , then the number of design variables will be $2l_h$. The complexity of this design problem should not be overlooked primarily for two reasons: first, since the PSNR depends on both the decimation filter H and interpolation filter G for a chosen encoder/decoder such as DCT-based transform coding scheme, it has no closed analytic expression; second, the constraint on bpp is difficult to formulate because it involves a coding scheme whose characterization is usually not purely analytic.

A. Objective Function

To formulate the design problem at hand as a minimization (rather than a maximization) problem, we propose to adopt the objective function

$$f(x) = -\text{PSNR} + w \cdot n_z$$

where x is a column vector that collects the $2l_h$ coefficients of $h_1(z_1)$ and $h_2(z_2)$, n_z denotes the normalized number of nonzero DCT coefficients after quantization and coding, and $w > 0$ is a weighting factor. Here the bpp is replaced by a more manageable n_z , and the constraint on bpp is eliminated by considering a weighted combination of $-\text{PSNR}$ and n_z so that we are now dealing with an *unconstrained* optimization problem instead. For a given x (i.e. a given decimation filter H), the evaluation of the above objective function involves a least squares design of interpolation filter G (so that the

PSNR can be calculated) and, along the way, computing the normalized number of nonzero DCT coefficients when the encoding is carried out (see Fig. 1).

B. Minimization Algorithm

The minimization of $f(x)$ for a given weight w is carried out using a quasi-Newton algorithm called the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm which is known for its robust performance and computational efficiency [12]. The algorithm does not require an explicit evaluation of the Hessian matrix, yet it is a descent algorithm with a convergent rate comparable with that of a Newton algorithm [12]. With a given initial point x_0 , in the k -th iteration the BFGS algorithm updates point x_k to $x_{k+1} = x_k + \alpha_k d_k$ where direction d_k is determined by $d_k = -S_k \cdot g_k$, g_k is the gradient of the objective function at x_k , and S_k is a positive definite matrix with S_0 being the identity matrix and is updated as follows:

$$\begin{aligned} S_{k+1} &= S_k + C_k \\ C_k &= \left(1 + \frac{\gamma_k^T S_k \gamma_k}{\gamma_k^T \delta_k} \right) \frac{\delta_k \delta_k^T}{\gamma_k^T \delta_k} - \frac{\delta_k \gamma_k^T S_k + S_k \gamma_k \delta_k^T}{\gamma_k^T \delta_k} \\ \delta_k &= x_{k+1} - x_k \\ \gamma_k &= g_{k+1} - g_k \end{aligned}$$

The iteration continues until the 2-norm of the difference between x_k and x_{k+1} is less than a prescribed tolerance, or the number of iterations exceeds a given upper bound.

V. SIMULATION RESULTS

The proposed design algorithm was applied to several test images including *barbara*, *boat* and *goldhill*, all of size 512×512 . Due to the space limitation, below only simulation results concerning image *Barbara* are reported, although similar results are observed for images *boat* and *goldhill*. The algorithm starts with an initial point x_0 obtained by applying a modified TEMG algorithm where a symmetric extension of image \hat{Y} was used and both the filters $h_1(z_1)$ and $h_2(z_2)$ were linear-phase lowpass FIR filters generated by a Chebyshev window of length 7 with normalized cutoff frequency $f_c = 0.9$. The encoding-decoding part of the system was performed by DCT-based JPEG. A total of fourteen designs were carried out with the weighting factor w varying from 15 to 34 so as to keep the bpp in the range from 0.07 to 0.19. Accordingly, the quality level for quantization matrix, which is a required parameter for JPEG, was set in the range between 8 and 25. In general for a lower quality level, a large value of w was used because the normalized number of nonzero DCT coefficients is fairly small. Thus a relatively large weighting factor is needed for significant presence of the term $w \cdot n_z$ in the objective function $f(x)$. Typically, the BFGS algorithm gives a satisfactory design with a small number of iterations ranging from one iteration to twelve iterations. The system

performance in terms of PSNR versus bpp is depicted in Fig. 4, where the solid line is for the system with an optimal decimation filter H as described in Section IV and an optimal interpolation filter G designed using the method of [6] with symmetric-extended \hat{Y} ; the dashed line is for the system as described in [6] with a Chebeshev-windowed based lowpass decimation filter H and an optimal interpolation filter G designed using the method of [6] with symmetric-extended \hat{Y} ; and the dash-dotted line is for a system similar to the preceding system but the symmetric extension was replaced by zero-padding. The overall PSNR improvement for the optimized system over that of [6] was found to be in the range from 0.2dB to 0.5 dB.

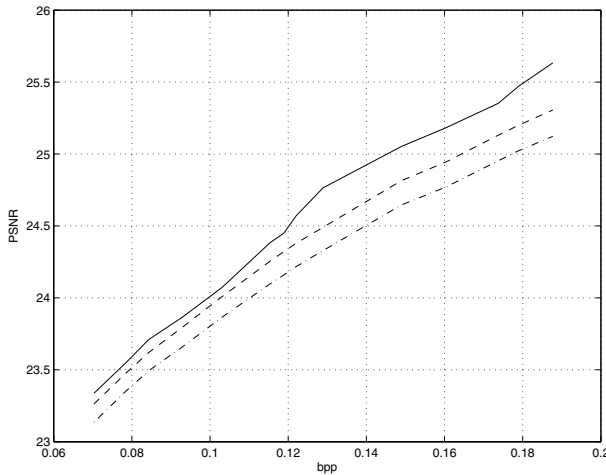


Fig. 4 Performance comparisons for image *barbara*.

Shown in Fig. 5 is the amplitude response of the optimized decimation filter obtained with quality level set to 12 and w set to 32. The image *barbara* produced by the system was found to have a PSNR = 24.7638 at bpp = 0.1295, which can be seen in Fig. 6(d). For comparison purposes, TEMG algorithm [6] and TEMG with symmetric extension were also applied at the same bpp and the PSNRs were found to be 24.3279 dB and 24.49 dB, respectively.

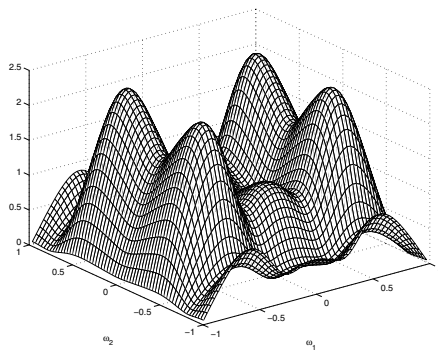


Fig. 5 Amplitude response of the optimal decimation filter H with quality level = 12 and $w = 32$.



(a) Original



(b) TEMG: PSNR = 24.3279dB



(c) Symm-Extension: PSNR = 24.4900dB



(d) Optimal: PSNR = 24.7638dB

Fig. 6 Image *barbara* at bpp = 0.1295 that are obtained using (b) TEMG algorithm, (c) TEMG algorithm with symmetric extension, and (d) optimal decimation filter combined with symmetric extension.

REFERENCES

- [1] "JPEG 2000 Image Coding System WG1N390 REV," ISO/IEC JTC1/SC 29/WG1.
- [2] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Compression Standard*, Van Nostrand Reinhold, New York, 1992.
- [3] "Digital Compression and Coding of Continuous-Tone Still Pictures, Part I, Requirements and Guidelines," ISO/IEC JTC1 Committee Draft 10918-1 Std., Feb. 1991.
- [4] "Information Technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 mbits," ISO/IEC 11172 Std., 1993.
- [5] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down-scaling for better transform compression," *IEEE Trans. on Image Processing*, vol. 12, no. 9, pp. 1132-1144, Sept. 2003.
- [6] Y. Tsaig, M. Elad, P. Milanfar, and G. H. Golub, "Variable projection for near-optimal filtering in low bit-rate block coders," *IEEE Trans. on Circuits Syst. Video Technology*, vol. 15, no. 1, pp. 154-160, Jan. 2005.
- [7] B. Zeng and A. N. Venetsanopoulos, "A JPEG-based interpolative image coding scheme," in *Proc. ICASSP*, vol. V, pp. 393-396, Minneapolis, MN., April 1993.
- [8] S. H. Jung, S. K. Mitra, and D. Mukherjee, "Subband DCT: Definition, analysis, and applications," *IEEE Trans. Circuits Syst. Video Technology*, vol. 6, no. 6, pp. 273-286, June 1996.
- [9] D. Mukherjee and S. K. Mitra, "Image resizing in the compressed domain using subband DCT," *IEEE Trans. Circuits Syst. Video Technology*, vol. 6, no. 7, pp. 620-627, July 1996.
- [10] R. Dugad and N. Ahuja, "A scheme for spatial scalability using non-scalable encoders," *IEEE Trans. Circuits Syst. Video Technology*, vol. 13, no. 10, pp. 993-999, Oct. 2003.
- [11] R. Gonzalez and R. Woods, *Digital Image Processing*, 2nd ed., Prentice Hall, 2002.
- [12] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Reading, MA., Wesley, 1984.