Perceptual dithering for octave subband image coding[☆]

Chung-Neng Wang,[1] Chi-Min Liu,[*,2] and Tihao Chiang[1]

*Department and Institute of Computer Science and Information Engineering,
National Chiao Tung University (NCTU), Hsinchu 30050, Taiwan*

Abstract

Octave subband coding has been used as an efficient image coding approach for various well-known standards such as MPEG-4 and JPEG-2000. Although the sub-images in the octave subband representation have low statistical correlation, the visual similarity among the sub-images indicates that there exists some unexploited perceptual redundancy between sibling bands in addition to the parent–child redundancy. In this paper, we show that the original image can be perceptually dithered to form a visually equivalent image with increased interband correlation that can be used to achieve more compression. To remove the redundancy among the sibling subbands of an image, we provide a novel perceptual dithering approach that is based on entropy reduction technique with perceptual model. The theoretical basis for the entropy reduction is proven by a theorem for Gaussian distributed signals. The use of our dithering technique is demonstrated for an MPEG-4 compliant encoder with improved coding efficiency using MPEG-4 Visual Texture Coding (VTC) based on zerotree entropy coding. Our results show that there exists a perceptual interband redundancy even though the original interband correlation is small. For a perceptually transparent image quality, our approach can achieve bit savings over MPEG-4 VTC by 11–30% while maintaining compatibility with the MPEG-4 standards.
© 2003 Elsevier Inc. All rights reserved.

[*] Corresponding author. Fax: +886-3-5724176.
E-mail addresses: cmliu@csie.nctue.du.tw, eldon@intervideo.com.tw (C.-M. Liu).
[1] Tihao Chiang is currently an Associate Professor and Chung-Neng Wang is a research fellow of the Department and Institute of Electronics Engineering, NCTU, Taiwan.
[2] Chi-Min Liu is currently a Professor of the Department and Institute of Computer Science and Information Engineering, NCTU, Taiwan.

1. Introduction

The fundamental goal of an image coding technique is to reduce the bitrate required for transmission and storage while maintaining subjective fidelity. This goal can be achieved by the removal of both the statistical and perceptual redundancy. Since octave subband coding can decompose a two-dimensional (2-D) image signal into several statistically less correlated subband signals that can match the human visual perceptual resolution, octave decomposition is now commonly used as a tool for efficient image coding (Antonini et al., 1992; Johnsen et al., 1990; Mallat, 1989; Mohsenian and Nasrabadi, 1994; MPEG, 2000; Shapiro, 1993; Vaisey, 1995; Woods and O'Neil, 1986). As shown in Fig. 1a the typical octave subbands are grouped into three octave levels L_0 to L_2 or 17 subbands B_0 to B_6. Fig. 1b shows the typical representation of the octave subbands for an image.

For an image, the relationships between octave subbands can be classified into two categories including the parent–child and the sibling subband relationships. Since the statistical correlation across octave subbands is generally much lower, distinct quantization schemes are designed for each subband tailoring to the statistical and perceptual characteristics of each subband (Antonini et al., 1992; Mallat, 1989; Woods and O'Neil, 1986). However, as observed from Fig. 1b, visually similar patterns in different subbands indicate that there is strong correlation among the wavelet coefficients of different subbands.
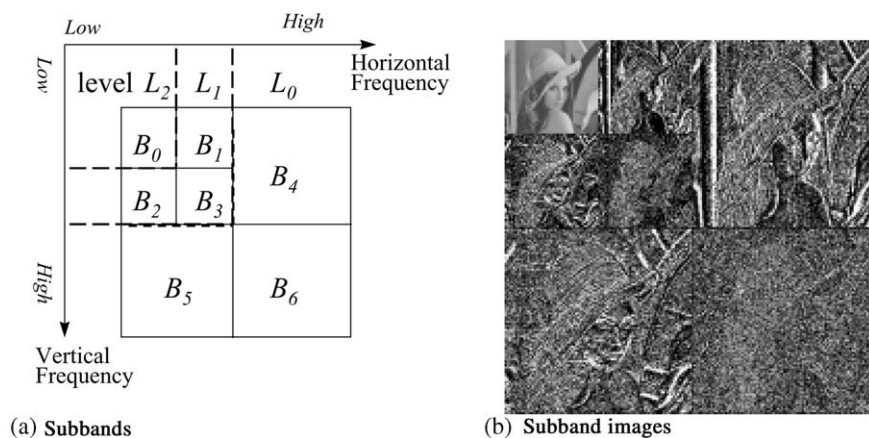

Fig. 1. A 2-D image is logarithmically partitioned into subbands or wavelet trees using the DWT. The symbol B_k denotes the kth subband and the symbol L_k denotes the kth level of the wavelet coefficients.

As for the parent–child relationship, there are two findings that are used to enhance the coding efficiency. The first finding is the zerotree structure of the subbands as noted by Shapiro in (Shapiro, 1993). The zerotree symbols and the range of the magnitude for the wavelet coefficients can be predicted in a low-to-high frequency subband levels (for example, from B_1 to B_4 in Fig. 1), and hence the embedded zerotree (EZW) scheme was developed to efficiently encode these zero symbols and coefficients. The second finding is the use of adaptive quantizers for individual images (Johnsen et al., 1990). Since there are similar patterns in different subbands, Johnsen et al. (1990) found that the image contents of the subbands at the higher level can be used to predict the characteristics of the image contents at the lower level subbands. The characteristics of the content can be classified into finite classes, and the quantizers with different ranges and step sizes can be designed and selected adaptively to match the statistics of the subband.

As for the sibling relationship, the key idea is to remove the redundancy among the sibling nodes of each wavelet tree through preprocessing. For example, the edge-based vector quantization (EBVQ) (Mohsenian and Nasrabadi, 1994) was proposed to provide better compression using vector quantization for subbands at the same level. The EBVQ technique requires edge-based codebooks and training session before the encoding process. It also requires decoder modification before the EBVQ approach can be used with a wavelet-based coding standard such as MPEG-4.

In order to remove the same redundancy while maintaining compatibility with MPEG-4 decoder, we propose a perceptual dithering coding approach (PDC) (Liu and Wang, 1998). The key innovation is that we can use perceptual dithering to replace the original image with a visually equivalent image by exploiting the redundancy among sibling nodes at the edge area where human visual system is less sensitive. This new technique will systemically modify the statistical distribution of the AC coefficients such that the entropy is reduced while the psychovisual fidelity is preserved (Chou and Li, 1995; Safranek and Johnston, 1989). Several images were tested to compare the proposed PDC and MPEG-4 Visual Texture Coding (VTC). Only the luminance component of color images and grayscale images were observed. While maintaining the visually lossless quality, we show that the PDC can achieve bitrate reduction for about 11–30%.

This paper is organized as follows. In Section 2, MPEG-4 VTC and the perceptual dithering technique are described. In Section 3, we will show the bitrate reduction results of the proposed PDC approach versus those of MPEG-4 VTC approach. The paper is concluded in Section 4.

2. Perceptual dithering coding

To exploit the sibling subband similarity, we propose a systematic PDC approach as shown in Fig. 2 to remove the redundancy among the siblings of the wavelet coefficients. In addition to the existing modules of MPEG-4 VTC, the PDC approach adds two more modules including the psychovisual model and the perceptual
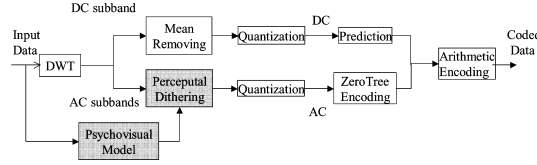
Fig. 2. Block diagram of the encoder for the perceptual dithering coding.

dithering. In the perceptual dithering module, the coefficients of the sibling nodes for each wavelet tree are perturbed according to the statistical distribution of AC coefficients within the subband at the same frequency level to achieve more energy compaction. The level of perturbation for each wavelet coefficient is constrained by the noise tolerance levels according to the psychovisual model. The resultant picture has the same visual quality but with lower entropy after dithering.

2.1. Rationale

As shown in Figs. 1a and b, an image is organized as logarithmically scaled discrete wavelet transform (DWT) subbands. The subband signals in the horizontal, the vertical, and the diagonal subbands are denoted as $s_h(m, n), s_v(m, n)$, and $s_d(m, n)$, respectively. Let a vector $S(m, n) = [s_h(m, n), s_v(m, n), s_d(m, n)]^T$ denote the three subband signals that are characterized by a real wide-sense stationary random process. Since we assumed the process is wide-sense stationary, the second-order statistic is sufficient to describe its behavior. The indices m and n are skipped for simplicity, i.e., $s_h(m, n), s_v(m, n), s_d(m, n)$, and $S(m, n)$ are abbreviated as s_h, s_v, s_d, and S. Because the octave subbands of higher frequency levels (for example, the subbands at the levels L_0 and L_1 in Fig. 1) have been shown to have (near) zero mean (Antonini et al., 1992; Shapiro, 1993), the autocorrelation and the autocovariance of the subbands are identical. The autocorrelation matrix $\mathbf{R}_S$ for the octave subbands at the same level has the following form:

$$\mathbf{R}_S = \begin{pmatrix} E(s_h \cdot s_h) & E(s_h \cdot s_v) & E(s_h \cdot s_d) \\ E(s_v \cdot s_h) & E(s_v \cdot s_v) & E(s_v \cdot s_d) \\ E(s_d \cdot s_h) & E(s_d \cdot s_v) & E(s_d \cdot s_d) \end{pmatrix}, \tag{1}$$

where the expectation function $E(\cdot)$ can be computed by the sample average of the wavelet coefficients $X(m, n)$ in the following:

$$E(X) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} X(m, n).$$

The values of the elements along the main diagonal denote the intraband correlation, while the values of off-diagonal elements denote the interband correlation. The interband correlation of the octave subbands at the same level can be found in matrix $\mathbf{R}_S$. For example, the matrix of the correlations between the octave subbands at level L_1 in Fig. 1b is computed as

$$\mathbf{R}_S = \begin{pmatrix} 115.66 & 11.27 & -0.17 \\ 11.27 & 54.88 & 0.08 \\ -0.17 & 0.08 & 17.47 \end{pmatrix}$$

$$= 110889.4 \times \begin{pmatrix} 1 & 0.1414 & -0.0038 \\ 0.1414 & 1 & 0.0026 \\ -0.0038 & 0.0026 & 1 \end{pmatrix} = 110889.4 \times \bar{\mathbf{R}}_S,$$

where the matrix $\bar{\mathbf{R}}_S$ has the normalized correlation coefficients.

With the matrix $\bar{\mathbf{R}}_S$, we can find that the values of off-diagonal elements, as compared to the values of the elements in the main diagonal, are significantly low. That is, the interband correlation is much lower than the intraband correlation for the subband signals at the same level.

There are similar patterns for the AC subbands which prompt the existence of the interband correlation. Thus, the correlations can be classified as the statistical interband correlation and the perceptual interband correlation. The statistical interband correlation tends to be low because of the DWT decorrelation property. However, the perceptual interband correlation can be exploited for more coding gain (Liu and Wang, 1998). In this paper, we provide a novel approach for reducing the remaining perceptual correlation. The other goal of the perceptual dithering technique is that the reconstructed picture quality is perceptually identical to the original image. Thus, the perceptual dithering facilitates efficient image coding without any quality loss in perception.

2.2. Perceptual model

For each wavelet coefficient, the perceptual model is constructed based on the visibility threshold, namely the pixel perceptual tolerable error (PPTE). In evaluating the visibility thresholds, the model assumes that experiments are conducted in a dark room with a 19-inch SPARC-workstation monitor and the viewing distance is six times the height of an image. The derivation of the PPTE follows the approach as described in (Chou and Li, 1995), which is based on both the background luminance perceptibility and the spatial masking effects. Our paper is not constrained by the development of more precise perceptual models.

Let the original image be denoted as the subband image at level -1. The full-band Just-Noticeable-Difference (JND) profile ($\mathrm{JND_{fb}}$) is initially assigned as the PPTE for the subband image at level -1. For a gray-level still image with size $H \times W$,

$$\mathrm{PPTE}_{0,-1}(m,n) = \mathrm{JND_{fb}}(m,n), \quad \text{for } 0 \leqslant m \leqslant H-1, \ 0 \leqslant n \leqslant W-1. \tag{2}$$

For the octave representation with L levels, the $\mathrm{PPTE}_{q,l}(m,n)$ at position (m,n) of the qth subband at level l is generated from the previous level in the following form:

$$\mathrm{PPTE}_{q,l+1}^2(m,n) = \left[\sum_{i=0}^{1} \sum_{j=0}^{1} \mathrm{PPTE}_{0,l}^2(i+2m, j+2n) \right] \cdot \varpi_{q,l}, \tag{3}$$

for $q = 0, \ldots, 3$, $l = 0, \ldots, L-1$, $0 \leqslant m < (H/2(l+1))$, and $0 \leqslant m < (W/2(l+1))$. The symbol $\varpi_{q,l}$ indicates the perceptual weight of the qth subband at level l. The perceptual weights (Chou and Li, 1995) are computed as the inverse of the normalized spatial-frequency sensitivity, which is defined as

$$\varpi_{q,l} = \left(\gamma_{q,l} \sum_{k=0}^{3} \gamma_{k,l}^{-1} \right)^{-1}, \quad \text{for } q = 0, \ldots, 3, \tag{4}$$

where $\gamma_{k,l}$ denotes, in terms of the signal energy, the average sensitivity of the human eyes to spatial frequencies in the kth subband at level l (Chou and Li, 1995). The average sensitivity $\gamma_{k,l}$ can be computed as follows:

$$\gamma_{k,l} = \frac{4(l+1)^2}{H \cdot W} \sum_{u=\varepsilon_k \cdot h}^{[(\varepsilon_k+1)h-1]} \sum_{v=\rho_k \cdot w}^{[(\rho_k+1)w-1]} \xi(u,v), \quad \text{for } k = 0, \ldots, 3, \tag{5}$$

where $h = (H/2(l+1))$, $w = (W/2(l+1))$, $\varepsilon_k = \lfloor k/2 \rfloor$, $\rho_k = k - \varepsilon_k \cdot 2$, and $\xi(u,v)$ is the sensitivity at the spatial frequency (u,v) calculated according to the modulation transfer function (MTF) response curve (Safranek and Johnston, 1989). The operation $\lfloor a \rfloor$ represents the largest integer that is less than or equals to a.

Thus, the PPTE of each wavelet coefficient is used to ensure the perceptually transparent coding of an image. Assuming the transparent coding is not achievable for some applications requiring low bitrate, coding approach using the minimally noticeable distortion (MND) is necessary. The visibility threshold MND is quantified as a multiple of the PPTE using the approaches as described in (Chou and Li, 1995).

2.3. Perceptual dithering

The dithering process compacts the energy of the wavelet coefficients from the sibling subbands to one chosen subband at the same frequency level. As shown in Fig. 3, we plot the vectors of three wavelet coefficients in a three-dimensional coordinate system. Assume the region occupied by the majority of the points lies close to the vector v_1 as illustrated in the figure, roughly the olive-shaped volume. Consider the data vector transformed by coordinate rotation to axes v_1, v_2, and v_3, which are referred to as the main eigenvectors of the three subbands at the same frequency level. In the new
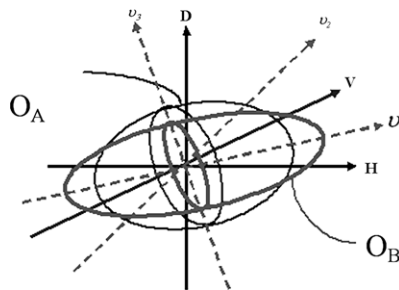

Fig. 3. Representation of coefficients distributions of three subband vectors in three dimensions.
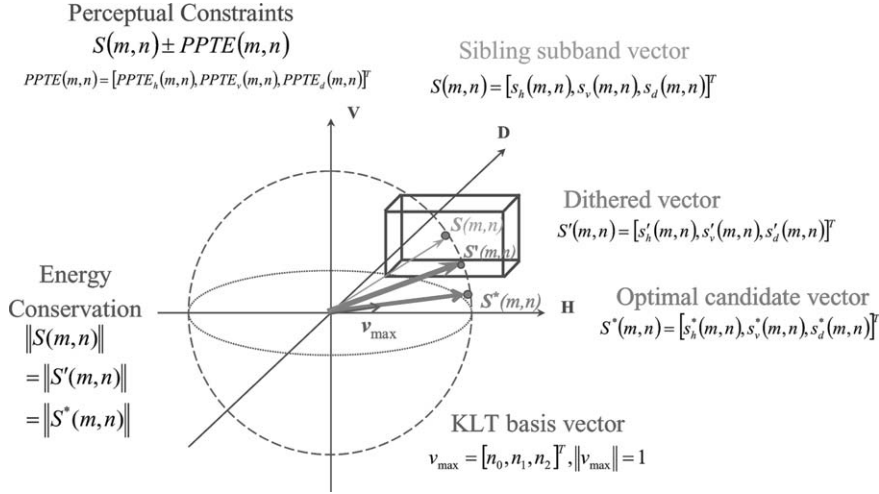
Fig. 4. Graphical illustration of the dithering along the eigenvector with maximum eigenvalue (Liu and Wang, 1998). Where H, V, and D denote the spatial orientations respectively. $v_{\max}$ is the target axis for energy compaction from the original vector $S(m,n)$ to the dithered vector $S'(m,n)$. The box constrains the permissible degree of rotation for the vector $S(m,n)$ without visible distortion.

coordinate system, the transformed coefficients all have v_1 components, which are much larger than those in the other components. Furthermore, we compact the energy of the coefficients in the v_2, and v_3 directions into that of coefficients in the v_1 direction by stretching the olive-shaped ball from O_A to O_B. Consequently, we find that the energy of the data vectors along the v_1 direction is increased and the energy of the data vectors in the other axes is decreased in this process, which is accomplished by perceptual dithering approach here. Thus, the eigenvalue along the v_1 axis is enlarged and those along with the v_2, and v_3 axes are decreased.

As shown in Fig. 4, the energy is concentrated by moving the vectors of the wavelet coefficients toward the target vector $v_{\max}$, which is the eigenvector with the maximum eigenvalue derived from the statistics of the sibling subbands for the specified frequency level. Most of the energy that is spread over the wavelet coefficients within the sibling subbands at the same frequency level will be concentrated at only a few wavelet coefficients. The associated visibility threshold is used to keep the dithering noise imperceptible. Thus, the perceptual dithering can compact the energy of the sibling subbands into the specified subband at the same frequency level with the same visual quality of picture.

The following theorem provides the principle for finding the dithered vector of $S(m,n)$.

Lemma 1. *Let X and X' be two Gaussian random vectors with zero mean and autocorrelation matrices $\mathbf{R}_X$ and $\mathbf{R}_{X'}$ respectively. Let the two sets of sorted eigenvalues for $\mathbf{R}_X$ and $\mathbf{R}_{X'}$ be $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ and $\{\lambda'_1, \lambda'_2, \cdots, \lambda'_N\}$, respectively, with equivalent energy, i.e.,*

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_N; \quad \lambda_1' \geqslant \lambda_2' \geqslant \cdots \geqslant \lambda_N', \tag{6}$$

and

$$\lambda_1 + \lambda_2 + \cdots + \lambda_N = \lambda_1' + \lambda_2' + \cdots + \lambda_N'. \tag{7}$$

If all the eigenvalues are identical except for two of them in each set, the random vector with the largest of the four eigenvalues has lower entropy.

Proof. Proof of this *Lemma* is given in Appendix A.

Theorem 1. *Let X and X' be two Gaussian random vectors with zero mean and autocorrelation matrices $\mathbf{R}_X$ and $\mathbf{R}_{X'}$ respectively. Let the two sets of sorted eigenvalues for $\mathbf{R}_X$ and $\mathbf{R}_{X'}$ be $\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$ and $\{\lambda_1', \lambda_2', \cdots, \lambda_N'\}$, respectively, with equivalent energy as described in Eqs. (6) and (7). If all the eigenvalues are identical except for the three of them in each set, the random vector with the largest and smallest values of the six eigenvalues has lowest entropy.*

Proof. Proof of the *Theorem* is listed in Appendix B.

The dithering process based on the primary eigenvector consists of three phases (Liu and Wang, 1998). First, the autocorrelation matrix for the subband images is computed with Eq. (1). Second, based on the autocorrelation matrix, the target vector $v_{\max}$ is evaluated by finding the eigenvector that has the maximum eigenvalue through Householder method and QL algorithm (Press et al., 1990). Third, each subband vector $S(m, n)$ is rotated toward the eigenvector $v_{\max}$ with a rotation angle constrained by the rectangular box as shown in Fig. 4. The third phase involves the derivation of the desired vector, which depends on the relative location between the box and the eigenvector. As a trivial example, when the primary eigenvector penetrates the box, the dithered vector is selected as the eigenvector $v_{\max}$. When the eigenvector does not penetrate the box, the desired vector is derived through the following steps.

For each vector from the three subbands at the lth level, the input vector $S(m, n)$ is dithered under the constraints of visibility threshold and the eigenvector. Since the dithering process is used for each level sequentially, the label of each level is skipped for simplicity.

The input vector is given by

$$S(m, n) = [s_{\mathrm{h}}(m, n), s_{\mathrm{v}}(m, n), s_{\mathrm{d}}(m, n)]^{\mathrm{T}}.$$

The perceptual thresholds are denoted as

$$\begin{aligned} \mathrm{PPTE}(m, n) &= [\mathrm{PPTE}_{\mathrm{h}}(m, n), \mathrm{PPTE}_{\mathrm{v}}(m, n), \mathrm{PPTE}_{\mathrm{d}}(m, n)]^{\mathrm{T}} \\ &= [\mathrm{PPTE}_{1,l}(m, n), \mathrm{PPTE}_{2,l}(m, n), \mathrm{PPTE}_{3,l}(m, n)]^{\mathrm{T}} \end{aligned}$$

and the eigenvector is denoted as

$$v_{\max} = [n_0, n_1, n_2]^{\mathrm{T}} \quad \text{with} \quad ||v_{\max}|| = 1.$$

The vector that has the identical energy as the input vector and points along the eigenvector, referred to as the on-axis vector, is defined as

$$S^*(m,n) = [s_h^*(m,n), s_v^*(m,n), s_d^*(m,n)]^{\mathrm{T}} = ||S(m,n)||v_{\max},$$

where $||X||$ indicates the energy of the random vector X with dimension N_X, represented by the sum of the squares of the entry values.

$$||X|| = \sum_{d=1}^{N_X} X_d^2. \tag{8}$$

After perceptual dithering, the output is referred to as the dithered vector as follows:

$$S'(m,n) = [s_h'(m,n), s_v'(m,n), s_d'(m,n)]^{\mathrm{T}}.$$

The perceptual dithering algorithm is summarized as follows:

Step 1. If the input vector has either nonzero energy or at least two nonzero visibility thresholds, compute the on-axis vector and derive the low and high bounds for every coefficient of the dithered vector.

Step 2. Check if the on-axis vector lies within the visibility thresholds. If this is true, set the on-axis vector as the dithered vector and stop the dithering process.

Step 3. Search all the candidates of the dithered vector that lies within the bounds of the visibility thresholds. Among the six candidates, select the candidate closest to the on-axis vector as the dithered vector and stop the dithering process.

In Step 1, compute the energy of the input vector $S(m,n)$ by Eq. (8). Since the dithering process has the unitary property, the dithered vector has identical energy as the input vector due to the energy conservation

$$||S(m,n)|| = ||S'(m,n)|| = ||S^*(m,n)||.$$

Additionally, the low and high bounds of the dithered coefficient at the specified subband within lth level can be found by $\mathrm{LB}_i(m,n) = s_i(m,n) - \mathrm{PPTE}_{i,l}(m,n)$ and $\mathrm{HB}_i(m,n) = s_i(m,n) + \mathrm{PPTE}_{i,l}(m,n)$ for $i \in \{h,v,d\}$, respectively.

Since the magnitude of each coefficient of the input vector is altered in Step 2 under the constraint of perceptual error tolerance, the dithered coefficients shall satisfy the following inequalities, which are:

$$\mathrm{LB}_i(m,n) \leqslant s_i'(m,n) \leqslant \mathrm{HB}_i(m,n), \quad \text{for every } i \in \{h,v,d\}. \tag{9}$$

The inequalities also provide the minimum and maximum magnitude for every dithered coefficient

$$s_j'(m,n) = \beta_j(m,n),$$

where $\beta_j(m,n) \in \{\mathrm{LB}_j(m,n), \mathrm{HB}_j(m,n)\}$ for $j \in \{h,v,d\}$.

If each coefficient of the on-axis vector satisfies the conditions as defined in Eq. (9), $\mathrm{LB}_i(m,n) \leqslant s_i^*(m,n) \leqslant \mathrm{HB}_i(m,n)$, we can claim that the dithered vector equals to the on-axis vector

$$S'(m,n) = S^*(m,n).$$

If the on-axis vector found in Step 3 is not the desired solution, we compute the dithered vector by exhaustively searching within the bounds for each coefficient. Geometrically, such an operation is identical to the search for the intersection of both the curve that is originated from the input vector to the on-axis vector and the specified plane, which contains the dithered vector, input vector, on-axis vector and the center of the ball in Fig. 4 simultaneously. Consequently, the perceptual dithering process is to compute the solution of the following sets of simultaneous equations:

$$||S(m,n)|| = ||S'(m,n)||,$$

$$\begin{aligned} s_h(m,n)^2 + s_v(m,n)^2 + s_d(m,n)^2 \\ = s'_h(m,n)^2 + s'_v(m,n)^2 + s'_d(m,n)^2 \end{aligned} \tag{10}$$

$$s_h(m,n) + a(m,n)s_v(m,n) + b(m,n)s_d(m,n) = 0, \tag{11}$$

$$s'_h(m,n) + a(m,n)s'_v(m,n) + b(m,n)s'_d(m,n) = 0, \tag{12}$$

$$\begin{aligned} s_h^*(m,n) + a(m,n)s_v^*(m,n) + b(m,n)s_d^*(m,n) \\ = n_0 + a(m,n)n_1 + b(m,n)n_2 = 0, \end{aligned} \tag{13}$$

where the variables $a(m,n)$ and $b(m,n)$ are the parameters for a plane. Thus, the remaining issue is focused on how to derive the values of the variables $a(m,n)$ and $b(m,n)$. To simplify the problem, let $s'_h(m,n) = \mathrm{LB}_h(m,n)$, Eqs. (10) and (12) become

$$s_h(m,n)^2 + s_v(m,n)^2 + s_d(m,n)^2 = \mathrm{LB}_h(m,n)^2 + s'_v(m,n)^2 + s'_d(m,n)^2, \tag{14}$$

and

$$\mathrm{LB}_h(m,n) + a(m,n)s'_v(m,n) + b(m,n)s'_d(m,n) = 0, \tag{15}$$

respectively.

By subtracting Eq. (13) from Eq. (11), we have

$$a(m,n) = -b(m,n)\frac{(s_d(m,n) - n_2)}{(s_v(m,n) - n_1)} - \frac{(s_h(m,n) - n_0)}{(s_v(m,n) - n_1)}. \tag{16}$$

Substitute the variable $a(m,n)$ at Eq. (16) into Eq. (15), we obtain

$$b(m,n) = \frac{s_h(m,n)(s_v(m,n) - n_1) + (s_h(m,n) - n_0)s_v(m,n)}{(s_d(m,n) - n_2)s_v(m,n) - s_d(m,n)(s_v(m,n) - n_1)}. \tag{17}$$

Substitute $b(m,n)$ back into Eq. (16), the variable $a(m,n)$ becomes

$$\begin{aligned} a(m,n) = &-\frac{s_h(m,n)(s_v(m,n) - n_1) + (s_h(m,n) - n_0)s_v(m,n)}{(s_d(m,n) - n_2)s_v(m,n) - s_d(m,n)(s_v(m,n) - n_1)} \\ &\times \frac{(s_d(m,n) - n_2)}{(s_v(m,n) - n_1)} - \frac{(s_h(m,n) - n_0)}{(s_v(m,n) - n_1)} \end{aligned} \tag{18}$$

With the derived $a(m,n)$ and $b(m,n)$, Eq. (12) becomes

$$
\begin{aligned}
s'_v(m,n) &= -\frac{b(m,n)}{a(m,n)} s'_d(m,n) - \frac{1}{a(m,n)} s'_h(m,n) \\
&= -\frac{b(m,n)}{a(m,n)} s'_d(m,n) - \frac{1}{a(m,n)} \mathrm{LB}_h(m,n).
\end{aligned}
\tag{19}
$$

Substitute $s'_v(m,n)$ into Eq. (14), the last coefficient $s'_d(m,n)$ can be solved by:

$$
s'_v(m,n)^2 + s'_d(m,n)^2 - \|s(m,n)\| + \mathrm{LB}_h(m,n)^2 = 0.
$$

$$
\left[\left(\frac{b(m,n)}{a(m,n)}\right)^2 + 1 \right] s'_d(m,n)^2 - 2 \times \frac{1}{a(m,n)} \times \frac{b(m,n)}{a(m,n)} s'_d(m,n) \mathrm{LB}_h(m,n)
$$
$$
- \left[\|(m,n)\| - \left(\left(\frac{1}{a(m,n)}\right)^2 + 1 \right) \mathrm{LB}_h(m,n)^2 \right] = 0,
\tag{20}
$$

Let

$$
\xi = \left[\left(\frac{b(m,n)}{a(m,n)}\right)^2 + 1 \right],
\tag{21}
$$

$$
\psi = -2 \times \frac{1}{a(m,n)} \times \frac{b(m,n)}{a(m,n)} \mathrm{LB}_h(m,n),
\tag{22}
$$

and

$$
\zeta = -\left[\|s(m,n)\| - \left(\left(\frac{1}{a(m,n)}\right)^2 + 1 \right) \mathrm{LB}_h(m,n)^2 \right].
\tag{23}
$$

By substituting ξ, ψ, and ζ, Eq. (20) can be simplified as

$$
\xi s'_d(m,n)^2 + \psi s'_d(m,n) + \zeta = 0.
\tag{24}
$$

Consequently, $s'_d(m,n)$ can be computed as the roots of the quadratic equation

$$
s'_d(m,n) = \frac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi}.
\tag{25}
$$

We can substitute the solved $s'_d(m,n)$ back to Eq. (19) to compute $s'_v(m,n)$. Through the similar procedure, we can find all possible candidates of the dithered vector within the bounds defined in Eq. (9). Among all candidates, the dithered vector is selected as the vector closest to the on-axis vector.

    In practice, we need to consider the magnitudes of the input vector and the eigen-vector in deriving the values of dithered vector. For example, let $s'_h(m,n) = \mathrm{LB}_h(m,n)$, $s'_d(m,n)$ and $s'_v(m,n)$ are derived by the following expressions:

$$
s'_h(m,n) = \mathrm{LB}_h(m,n),
\tag{26}
$$

(a) If $n_1 = 0$ and $n_2 = 0$, $s'_{\mathrm{d}}(m,n)$ and $s'_{\mathrm{v}}(m,n)$ equal to

$$
s'_{\mathrm{v}}(m,n) = \begin{cases}
0, & s_{\mathrm{v}}(m,n) = 0, \\
\mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2}, & s_{\mathrm{v}}(m,n) \neq 0, \\
& s_{\mathrm{d}}(m,n) = 0, \\
\mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{\dfrac{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2}{[s_{\mathrm{d}}(m,n) \times s_{\mathrm{v}}(m,n)^{-1}]^2 + 1}}, & s_{\mathrm{v}}(m,n) \neq 0, \\
& s_{\mathrm{d}}(m,n) \neq 0,
\end{cases}
\tag{27}
$$

$$
s'_{\mathrm{d}}(m,n) = \begin{cases}
0, & s_{\mathrm{d}}(m,n) = 0, \\
\mathrm{sign}(s_{\mathrm{d}}(m,n)) \times \sqrt{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2}, & s_{\mathrm{v}}(m,n) = 0, \\
& s_{\mathrm{d}}(m,n) \neq 0, \\
s_{\mathrm{d}}(m,n) \times s_{\mathrm{v}}(m,n)^{-1} \times s'_{\mathrm{v}}(m,n), & s_{\mathrm{v}}(m,n) \neq 0, \\
& s_{\mathrm{d}}(m,n) \neq 0.
\end{cases}
\tag{28}
$$

(b) If $n_1 = 0$ and $n_2 \neq 0$, $s'_{\mathrm{d}}(m,n)$ and $s'_{\mathrm{v}}(m,n)$ are

$$
s'_{\mathrm{v}}(m,n) = \begin{cases}
0, & s_{\mathrm{v}}(m,n) = 0, \\
\mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{d}}(m,n)^2}, & s_{\mathrm{v}}(m,n) \neq 0,
\end{cases}
\tag{29}
$$

$$
s'_{\mathrm{d}}(m,n) = \begin{cases}
\mathrm{sign}(s_{\mathrm{d}}(m,n)) \times \sqrt{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2}, & s_{\mathrm{v}}(m,n) = 0, \\
\dfrac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi} \text{ by Eq.(25)}, & s_{\mathrm{v}}(m,n) \neq 0, \\
& s_{\mathrm{d}}(m,n) \neq 0, \\
\dfrac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi} \text{ by Eq.(25)}, & s_{\mathrm{h}}(m,n) \neq 0, \\
& s_{\mathrm{v}}(m,n) \neq 0, \\
& s_{\mathrm{d}}(m,n) = 0, \\
(n_1 n_0^{-1}) \times s'_{\mathrm{h}}(m,n), & s_{\mathrm{h}}(m,n) = 0, \\
& s_{\mathrm{v}}(m,n) \neq 0, \\
& s_{\mathrm{d}}(m,n) = 0.
\end{cases}
\tag{30}
$$

(c) If $n_1 \neq 0$ and $n_2 = 0$, $s'_{\mathrm{d}}(m,n)$ and $s'_{\mathrm{v}}(m,n)$ become

$$
s'_{\mathrm{v}}(m,n) = \begin{cases}
\mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2}, & s_{\mathrm{d}}(m,n) = 0, \\
\mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{d}}(m,n)^2}, & s_{\mathrm{h}}(m,n) \neq 0, \\
& s_{\mathrm{v}}(m,n) = 0, \\
& s_{\mathrm{d}}(m,n) \neq 0, \\
(n_1 n_0^{-1}) \times s'_{\mathrm{h}}(m,n), & s_{\mathrm{h}}(m,n) = 0, \\
& s_{\mathrm{v}}(m,n) = 0, \\
& s_{\mathrm{d}}(m,n) \neq 0, \\
\mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{\|s(m,n)\| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{d}}(m,n)^2}, & s_{\mathrm{v}}(m,n) \neq 0, \\
& s_{\mathrm{d}}(m,n) \neq 0,
\end{cases}
\tag{31}
$$

$$s'_{\mathrm{d}}(m,n) = \begin{cases} 0, & s_{\mathrm{d}}(m,n) = 0, \\[2mm] \frac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi} \text{ by Eq. (25)}, & \begin{array}{l} s_{\mathrm{h}}(m,n) \neq 0, \\ s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) \neq 0, \end{array} \\[4mm] \mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{||s(m,n)|| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{v}}(m,n)^2}, & \begin{array}{l} s_{\mathrm{h}}(m,n) = 0, \\ s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) \neq 0, \end{array} \\[4mm] \frac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi} \text{ by Eq. (25)} & \begin{array}{l} s_{\mathrm{v}}(m,n) \neq 0, \\ s_{\mathrm{d}}(m,n) \neq 0. \end{array} \end{cases}$$

$$(32)$$

(d) If $n_1 \neq 0$ and $n_2 \neq 0$, we get $s'_{\mathrm{d}}(m,n)$ and $s'_{\mathrm{v}}(m,n)$ through

$$s'_{\mathrm{v}}(m,n) = \begin{cases} \mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{\frac{||s(m,n)|| - s'_{\mathrm{h}}(m,n)^2}{(n_2 n_1^{-1})^2 + 1}}, & \begin{array}{l} s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) = 0, \end{array} \\[4mm] \mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{||s(m,n)|| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{d}}(m,n)^2}, & \begin{array}{l} s_{\mathrm{h}}(m,n) \neq 0, \\ s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) \neq 0, \end{array} \\[4mm] \mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{||s(m,n)|| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{d}}(m,n)^2}, & \begin{array}{l} s_{\mathrm{h}}(m,n) = 0, \\ s_{\mathrm{v}}(m,n) \neq 0, \\ s_{\mathrm{d}}(m,n) = 0, \end{array} \\[4mm] \mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{||s(m,n)|| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{d}}(m,n)^2}, & \begin{array}{l} s_{\mathrm{v}}(m,n) \neq 0, \\ s_{\mathrm{d}}(m,n) \neq 0, \end{array} \\[4mm] (n_1 n_0^{-1}) \times s'_{\mathrm{h}}(m,n), & \begin{array}{l} s_{\mathrm{h}}(m,n) = 0, \\ s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) \neq 0, \end{array} \end{cases}$$

$$(33)$$

$$s'_{\mathrm{d}}(m,n) = \begin{cases} (n_2 n_0^{-1}) \times s'_{\mathrm{v}}(m,n), & \begin{array}{l} s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) = 0, \end{array} \\[4mm] (n_2 n_0^{-1}) \times S'_{\mathrm{h}}(m,n), & \begin{array}{l} s_{\mathrm{h}}(m,n) = 0, \\ s_{\mathrm{v}}(m,n) \neq 0, \\ s_{\mathrm{d}}(m,n) = 0, \end{array} \\[4mm] \frac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi} \text{ by Eq. (25)}, & \begin{array}{l} s_{\mathrm{h}}(m,n) \neq 0, \\ s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) \neq 0, \end{array} \\[4mm] \frac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi} \text{ by Eq. (25)}, & \begin{array}{l} s_{\mathrm{h}}(m,n) \neq 0, \\ s_{\mathrm{v}}(m,n) \neq 0, \\ s_{\mathrm{d}}(m,n) = 0, \end{array} \\[4mm] \frac{-\psi \pm \sqrt{\psi^2 - 4\xi\zeta}}{2\xi} \text{ by Eq. (25)}, & \begin{array}{l} s_{\mathrm{v}}(m,n) \neq 0, \\ s_{\mathrm{d}}(m,n) \neq 0, \end{array} \\[4mm] \mathrm{sign}(s_{\mathrm{v}}(m,n)) \times \sqrt{||s(m,n)|| - s'_{\mathrm{h}}(m,n)^2 - s'_{\mathrm{v}}(m,n)^2}, & \begin{array}{l} s_{\mathrm{h}}(m,n) = 0, \\ s_{\mathrm{v}}(m,n) = 0, \\ s_{\mathrm{d}}(m,n) \neq 0. \end{array} \end{cases}$$

$$(34)$$

The same derivation procedures from Eqs. (26)–(34) are sequentially adopted to check each boundary condition defined in Eq. (9). Consequently, the dithered vector that has the minimum distance away from the on-axis vector is selected as the output of the perceptual dithering process.

Note that the energy of the vector $S(m, n)$ is conserved during the dithering process. The energy is compacted through this dithering process so that the entropy of the resultant sibling subbands is minimized. Assume that the random vector S of the original wavelet coefficients is dithered and subsequently decorrelated into the new random vector S_A, where the autocorrelation matrix is diagonalized.

Assume that the vector X is a Gaussian random vector with N Gaussian random variables, the rate-distortion function gives the bound for coding X (Berger, 1971) as

$$D(\theta) = \frac{1}{N} \sum_{k=1}^{N} \min(\theta, \sigma_k^2) \quad \text{and} \quad R_X(\theta) = \frac{1}{N} \sum_{k=1}^{N} \max\left(0, \frac{1}{2} \log_2 \frac{\sigma_k^2}{\theta}\right),$$

where θ is the specified mean distortion and $D(\theta)$ is the associated average distortion. When the distortion is small enough such that $0 < \theta < \min_k\{\sigma_k^2\}$, we have $D(\theta) = \theta$. The rate-distortion bound for coding X is computed as

$$R_X(\theta_X) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{2} \log_2 \frac{\sigma_k^2}{D(\theta_X)} = \frac{1}{2} \log_2 \left(\prod_{k=1}^{N} \sigma_k^2\right)^{\frac{1}{N}} - \frac{1}{2} \log_2 D(\theta_X),$$

where θ_X is the specified distortion for vector X. In the case of decorrelating vector X with KL transform, the variances are the eigenvalues of autocorrelation matrix $\mathbf{R}_X$, i.e., $\sigma_k^2 = \lambda_k$.

Consequently, for the same perceptual distortion $D(\theta) = D(\theta_S) = D(\theta_{S_A})$, the theoretically achievable bitrate reduction per coefficient can be computed as

$$G_{S \to S_A} = R_S(\theta_S) - R_{S_A}(\theta_{S_A}) = \frac{1}{2} \log_2 \left(\prod_{k=1}^{N} \sigma_k^2\right)^{\frac{1}{N}} - \frac{1}{2} \log_2 \left(\prod_{k=1}^{N} \lambda_k\right)^{\frac{1}{N}}, \tag{35}$$

where σ_k^2 is the diagonal elements of the autocorrelation matrix for S and λ_k is the diagonal elements of the autocorrelation matrix for S_A. Thus, for the same visual quality, the perceptually dithering approach reduces the bitrate with more energy compaction across the sibling nodes of every wavelet tree in the image.

2.4. Coding efficiency of perceptually dithering approach

The bitrate reduction per coefficient of the perceptual dithering approach is tested for various types of subband filters (Liu and Wang, 1998). The theoretical bitrate reduction through full decorrelation using Karhunen Loéve transform is less than 0.05 bits per coefficient. After the perceptual dithering, the bitrate reduction achieved is about 0.393 bits for coding of the L_0 sub-images of Barbara and is about 0.893 bits for coding of the L_0 sub-images of Lena, where the linear phase filter in (Kronander, 1989) was used. The overall entropy reductions for Barbara and Lena are 18.53% and 54.91%, respectively.

In the experimental results using the non-ideal filters as described in (Liu and Wang, 1998), the evaluation of the PPTE is influenced by various characteristics of the filters such as the linearity of the phase, the sidelobe attenuation, and the regularity.

(1) *Linear phase*. Since the linear phase filters have identical spatial delay for each band (Woods and O'Neil, 1986), an analysis based on the same location in the bands will be better aligned. Such conjecture is confirmed by the observation that the linear phase filter Kro_97 (Kronander, 1989) provides the best performance among all the other filters used in (Liu and Wang, 1998). However, theoretic analysis of the performance behavior is difficult to formulate because typical image has non-stationary statistical characteristics. Similar observation and analysis were found in (Andrew et al., 1994, 1995; da Silva and Ghanbari, 1996).

(2) *Sidelobe attenuation*. Since a larger sidelobe leaks more energy into other bands, which makes each subband contain information from the other bands. Consequently, the PPTEs that are derived from the original image of each subband are not adequate. Hence, the perceptual dithering using the filters with smaller sidelobe attenuation generally makes a smaller dithering gain. For example, the filters as described in (Daubechies, 1992) do not perform well due to less sidelobe attenuation of the filters.

(3) *Regularity*. Filters with regularity retain a comfortable image quality when coarse quantizers are used (Antonini et al., 1992). In other words, irregular filters typically amplify noise. The noise by perceptual dithering can be controlled with accurate perceptual model. Therefore, regularity is not critical for dithering as observed from the experiments on the irregular filters such as HV_12 and HV_16 (Vaidyanathan and Hoang, 1988).

2.5. Constraints for MPEG-4 visual texture coding

The encoder of MPEG-4 VTC includes four basic modules: DWT, quantization process, zerotree coding (ZTC), and arithmetic coding. The DWT coefficients are grouped into the DC band and the AC bands according to frequency levels and spatial orientations. The DWT decomposes an image into three orientations: horizontal, vertical, and diagonal directions. The DC band is encoded separately from the AC bands. After quantization, the DC band is DPCM encoded and the AC bands are encoded by ZTC and arithmetic coder. To prevent the coding distortion from being visible, the PDC should consider the errors caused by both the quantization and the perceptual dithering processes.

Let $\psi = [\psi_1, \ldots, \psi_M]^{\mathrm{T}}$ be the vector of the wavelet coefficients and $\tilde{\psi}$ be a scalar quantized version of ψ. For each integer n, the quantized coefficients can be computed with two methods as follows:

$$\tilde{\psi}_i = \begin{cases} (n + \frac{1}{2})\Delta_i, & \psi_i \in [n\Delta_i, (n+1)\Delta_i) \\ n\Delta_i, & \psi_i \in [(n - \frac{1}{2})\Delta_i, (n + \frac{1}{2})\Delta_i) \end{cases}, \tag{36}$$

for any set of positive, finite quantization step sizes Δ_i, $i = 1, \ldots, k$. In MPEG-4 VTC, the first quantization method in Eq. (36) is used to the quantized noise

introduced by MPEG-4 VTC is at most half of the quantization step size. Since the total noise comes from both the quantization and dithering processes, the impact of the quantization noise on the reconstructed quality should be considered in computing the practical threshold. In order to maintain identical fidelity, the threshold for perceptual dithering is equivalent to the difference between the PPTE and the weighted value of the maximum quantization noise. The effect of the rounding errors should be included when wavelet coefficients are rounded to an integer. Thus, we are only allowed to modify or dither the absolute value of each coefficient within the following threshold:

$$B_{q,l} \leqslant \text{PPTE}_{q,l} - \alpha \left(\frac{\Delta_l}{2} \right) - d_{\text{rounding}},$$

where the d_{rounding} equals to 1.0 here. The weighting value α indicates the percentage of the permissible quantization noise for moving the energy of each coefficient around while keeping the degradation imperceptible when finite quantization step sizes Δ_i are used.

3. Experimental results

Both the PDC and MPEG-4 VTC are tested for comparisons. Table 1 shows the rate-distortion performance of the PDC and MPEG-4 PDC techniques. To assess the visual quality of the reconstructed images and to evaluate the coding efficiency of the proposed PDC, a quality measure based on human visual system is used instead of the typical Peak signal-to-nose ratio (PSNR). Thus, the fidelity metric is used to measure the subjective quality of the reconstructed image. The subjective evaluation is performed in the wavelet domain, which can be transformed to the spatial domain for quality assessment (Chou and Li, 1995). The new criterion, namely peak signal-to-perceptual-noise for subbands (PSPNRsub) (Wang et al., 2001), is defined as

$$\text{PSPNRsub}_{q,l} = 10 \log_{10} \frac{255^2}{E[e_{q,l}^2]},\tag{37}$$

$$e_{q,l}(m,n) = \begin{cases} 0 & \text{if } |d_{q,l}(m,n)| \leqslant \text{PPTE}_{q,l}(m,n), \\ e'_{q,l}(m,n) & \text{Otherwise}, \end{cases}$$

where $d_{q,l}(m,n) = S_{q,l}(m,n) - S'_{q,l}(m,n)$ is the coding error and the $e'_{q,l}(m,n) = |d_{q,l}(m,n)| - \text{PPTE}_{q,l}(m,n)$ indicates the amount of the perceptible error. In our simulation using MPEG-4 VTC with PDC, the reconstructed images are perceptually lossless, which is confirmed by evaluating the quality of each reconstructed image with the criterion in Eq. (37) and the subjective evaluation on computer screen. As defined in MPEG-4 VTC, the DWT uses the default integer type filter with length of (9, 3). The single quantization scheme (MPEG, 2000) was used for all simulations.

Table 1
The performance of the PDC versus MPEG-4 VTC with various quantization stepsizes ($= \Delta$) and different weights ($= \alpha$) of quantization distortion for Akiyo and Lena

Akiyo	Δ	α	Bitrate			PSNR (dB)	Bit saving (%)
			DC	AC	Total		
(a) Akiyo (Y)							
MPEG4	1	1	9023	322144	331167	54.74	
PDC	1	1	9023	213534	222557	43.58	32.80
	1	0.9	9023	213319	222342	43.56	32.86
	1	0.5	9023	211372	220395	43.48	33.45
	1	0	9023	209917	218940	43.16	33.89
MPEG4	2	1	9023	187760	196783	49.50	
PDC	2	1	9023	149873	158896	43.43	19.25
	2	0	9023	145500	154523	42.79	21.48
MPEG4	3	1	9023	129073	138096	47.95	
PDC	3	1	9023	112205	121228	43.53	12.21
	3	0.67	9023	110974	119997	43.38	13.11
(b) Lena							
MPEG4	1	1	30352	1142611	1172963	54.73	
PDC	1	1	30352	985576	1015928	41.65	13.39
	1	0.9	30352	982731	1013083	41.62	13.63
	1	0.5	30352	969050	999402	41.47	14.80
	1	0	30352	962858	993210	40.97	15.32
MPEG4	2	1	30352	830837	861189	48.23	
PDC	2	1	30352	741244	771596	41.52	10.40
	2	0	30352	710948	741300	40.47	13.92
MPEG4	3	1	30352	638985	669337	45.87	
PDC	3	1	30352	570896	601248	41.51	10.17
	3	0.67	30352	564216	594568	41.30	11.17

The symbol 'MPEG4' indicates the performance of MPEG-4 VTC while the 'PDC' indicates the performance of the PDC. The performance of the proposed PDC will be analyzed based on several factors including the quantization step sizes, the weights for the quantization noise. We evaluated several grayscale images and color images in YUV format. For the color images, only the luminance component (Y) was processed by the PDC. For CIF (352×288) format, one image namely Akiyo is tested. For the grayscale format, one image of size 512×512, namely Lena, is illustrated in Fig. 5. For each image in Table 1 and Fig. 5, the visual quality is infinite in PSPNRsub and the visual quality in PSNR is listed.

For visually transparent coding of an image, a finer quantizer is used such that the PDC can achieve bitrate reductions for about 11–30% on the average. For a coarse quantizer, the gains on the average are about 1–6% in bitrate reduction. As the quantization step size is increased, the coding gain is reduced since most of the AC coefficients are zeroed out through the quantization process. In Table 1 we demonstrate the influences of the weighting values on the performance of the PDC. The performance

Fig. 5. The reconstructed image of size 512×512 and in grayscale format, namely Lena, using MPEG-4 VTC and the proposed perceptual dithering coding. The quantization step size (Δ) used is 2 and the two weighting values (α) are used for the perceptually dithering. It should be noted that these images have been verified to be visually indistinguishable from the original image under the same viewing environments where the psychovisual model is created as described in Section 2.

difference in coding gains for various weights indicates that an optimal weighting can be used to further improve the coding efficiency of the proposed PDC. In addition, our results show that the perceptual dithering can systemically modify the statistical distribution of the AC coefficients such that the entropy is reduced while the psycho-visual fidelity is preserved. The changed statistical distribution of the AC coefficients not only increases the probability of the zero symbols but also improves the performance of entropy coding and arithmetic coding in MPEG-4 VTC. Thus, the PDC can reduce the bit rate of MPEG-4 VTC for coding an image without quality loss in perception.

4. Conclusions

In this paper, we developed a perceptual dithering method to increase the interband correlation for octave subband coding. The perceptual dithering method exploits the perceptual redundancy existed among octave subbands and dithers the original subband image to reduce entropy. This observation is further formalized with our theoretical investigation based on the entropy reduction rotation procedure. The subbands at the same level are systematically rotated or dithered toward the eigenvector to reduce the entropy. Such a dithering facilitates the redundancy removal compression process. Our approach elegantly compact the energy such that no modifications at the decoder is necessary. We have examined our techniques for various types of filters. Our results show that the filter with linear phase and bigger sidelobe attenuation is a better choice for dithering, while regularity of the filter is not a significant factor.

We also presented an example using the PDC technique to improve MPEG-4 VTC for images with transparent quality. The performance of the PDC is analyzed based on two factors: the quantization step sizes and the weighting of the quantization noise. The improvement can range from 11 to 30%. The coding gains are less at lower bitrates due to coarse quantization. In conclusion, the PDC approach is an MPEG-4 compatible preprocessing technique that achieves significant improvement in coding efficiency by exploiting the perceptual correlation between the sibling subband nodes within every wavelet tree at the same frequency level for coding of images with transparent visual quality.

Appendix A. Proof of Lemma 1

If all the eigenvalues are the same except the two of them in each set, we can imply the following relationship:

$$\lambda_i = \lambda_i' \quad \text{for} \quad i \neq k, \ i \neq l, \ 1 \leqslant k < l \leqslant N \quad \text{and} \quad 1 \leqslant i \leqslant N. \tag{A.1}$$

Combining Eq. (A.1) with the constraint of $\sum_{i=1}^{N} \lambda_i = \sum_{i=1}^{N} \lambda_i'$ in Eq. (7) yields

$$\lambda_k + \lambda_l = \lambda_k' + \lambda_l', \quad \text{where } k < l. \tag{A.2}$$

Assuming, without loss of generality, that X' has the largest value of these four values produces $\lambda_k' > \lambda_k$. Since the eigenvalues in the sets are sorted as $\lambda_0 \geqslant \lambda_1 \geqslant \cdots \geqslant \lambda_{N-1}$ and $\lambda_0' \geqslant \lambda_1' \geqslant \cdots \geqslant \lambda_{N-1}'$, we have

$$\lambda_k' > \lambda_k \geqslant \lambda_l > \lambda_l'. \tag{A.3}$$

The bitrate bounds required to represent X and X' are computed as

$$R_X(\theta_X) = \frac{1}{2} \log_2 (\lambda_1 \lambda_2 \cdots \lambda_N)^{\frac{1}{N}} - \frac{1}{2} \log_2 D(\theta_X) \text{ bits}, \quad D(\theta_X) < \min_k \{\lambda_k\}, \tag{A.4}$$

and

$$R_X'(\theta_{X'}) = \frac{1}{2} \log_2 (\lambda_1' \lambda_2' \cdots \lambda_N')^{\frac{1}{N}} - \frac{1}{2} \log_2 D(\theta_{X'}) \text{ bits}, \quad D(\theta_{X'}) < \min_k \{\lambda_k'\}. \tag{A.5}$$

For the same distortion, the difference between the bitrates required for coding X and X' is $(R_X(\theta_X) - R_{X'}(\theta_{X'}))$ bits. To prove that X' has a bitrate less than X is equivalent to proving $(R_X(\theta_X) - R_{X'}(\theta_{X'})) > 0$. According to Eqs. (A.4) and (A.5), the inequality $(R_X(\theta_X) - R_{X'}(\theta_{X'})) > 0$ can be easily demonstrated by showing $\lambda_1 \lambda_2 \cdots \lambda_N > \lambda'_1 \lambda'_2 \cdots \lambda'_N$. Since the two sets of eigenvalues are the same except the values λ'_k, λ'_l, λ_k and λ_l, the inequality $\lambda_1 \lambda_2 \cdots \lambda_N > \lambda'_1 \lambda'_2 \cdots \lambda'_N$ is reduced to $\lambda_k \cdot \lambda_l > \lambda'_k \cdot \lambda'_l$. Let $d = \lambda'_k - \lambda_k$. From Eqs. (A.2) and (A.3), we have $d = \lambda'_k - \lambda_k = \lambda_l - \lambda'_l > 0$. Thus,

$$\lambda_k \cdot \lambda_l - \lambda'_k \cdot \lambda'_l = \lambda_k \cdot \lambda_l - (\lambda_k + d)(\lambda_l - d) = (\lambda_k - \lambda_l)d + d^2 > 0. \qquad \square \quad \text{(A.6)}$$

Appendix B. Proof of Theorem 1

The condition that all the eigenvalues are the same except the three of them in each set implies the following relationship:

$$\lambda_i = \lambda'_i \quad \text{for} \quad i \neq j, \; i \neq k, \; i \neq l, \; 1 \leqslant j < k < l \leqslant N \quad \text{and} \quad 1 \leqslant i \leqslant N. \quad \text{(B.1)}$$

Combining Eq. (B.1) with the condition $\sum_{i=1}^{N} \lambda_i = \sum_{i=1}^{N} \lambda'_i$ in Eq. (7) yields

$$\lambda_j + \lambda_k + \lambda_l = \lambda'_j + \lambda'_k + \lambda'_l, \quad \text{where } j < k < l. \quad \text{(B.2)}$$

Let random vector X' have the largest and the smallest values of these six eigenvalues, which implies that $\lambda'_j > \lambda_j$ and $\lambda'_l < \lambda_l$. Since $\lambda_0 \geqslant \lambda_1 \geqslant \cdots \geqslant \lambda_{N-1}$ and $\lambda'_0 \geqslant \lambda'_1 \geqslant \cdots \geqslant \lambda'_{N-1}$, all the eigenvalues except λ'_k have the relationship: $\lambda'_j > \lambda_j \geqslant \lambda_k \geqslant \lambda_l > \lambda'_l$.

Similar to the proof in Lemma 1, we will prove that the bitrate for coding X' is less than that for X if we can show $\lambda_j \cdot \lambda_k \cdot \lambda_l > \lambda'_j \cdot \lambda'_k \cdot \lambda'_l$. Let $d_j = \lambda'_j - \lambda_j, d_l = \lambda_l - \lambda'_l$, and $d_k = \lambda'_k - \lambda_k$. From Eqs. (6) and (B.2), we get $d_j > 0, d_l > 0$, and $d_k = d_l - d_j$.

The following three cases are exhaustively checked:

Case 1: $d_k = \lambda'_k - \lambda_k = d_l - d_j > 0$.

Following the same derivation method in Eq. (A.6) yields $\lambda_j \cdot \lambda_k \cdot \lambda_l > (\lambda_j + d_j) \cdot \lambda_k \cdot (\lambda_l - d_j)$.

The same derivation method leads to $\lambda_j \cdot \lambda_k \cdot \lambda_l > (\lambda_j + d_j) \cdot (\lambda_k + d_k) \cdot (\lambda_l - d_j - d_k) = \lambda'_j \cdot \lambda'_k \cdot \lambda'_l$.

Case 2: $d_k = \lambda'_k - \lambda_k = 0$.

Since $\lambda'_k = \lambda_k$, the inequality $\lambda_j \cdot \lambda_k \cdot \lambda_l > \lambda'_j \cdot \lambda'_k \cdot \lambda'_l$ becomes $\lambda_j \cdot \lambda_l > \lambda'_j \cdot \lambda'_l$, which can be proved by invoking Lemma 1.

Case 3: $d_k = \lambda'_k - \lambda_k = d_l - d_j < 0$.

Following the same derivation procedure as that in Case 1, we get

$$\lambda_j \cdot \lambda_k \cdot \lambda_l > (\lambda_j + d_l) \cdot \lambda_k \cdot (\lambda_l - d_l) > (\lambda_j + d_l - d_k) \cdot (\lambda_k + d_k) \cdot (\lambda_l - d_l)$$
$$= \lambda'_j \cdot \lambda'_k \cdot \lambda'_l. \qquad \square$$

References

Andrew, J.P., Ogunbona, P.O., Paoloni, F.J., 1994. Coding gain and spatial localisation properties of discrete wavelet transform filter banks for image coding. Proc. IEEE Int. Conf. Image Process. 3 (13–16), 348–352.

Andrew, J.P., Ogunbona, P.O., Paoloni, F.J., 1995. Coding gain and spatial localisation properties of discrete wavelet transform filter banks for image coding. IEE Proc. Vision, Image Signal Process. 142 (3), 133–140.

Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I., 1992. Image coding using wavelet transform. IEEE Trans. Image Process. 1 (2), 205–220.

Berger, T., 1971. Rate Distortion Theory: A Mathematical Basis for Data Compression. Englewood Cliffs, NJ: Prentice-Hall.

Chou, C.H., Li, Y.C., 1995. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. IEEE Trans. Circuits Syst. Video Technol. 5 (6), 467–476.

da Silva, E.A.B., Ghanbari, M., 1996. On the performance of linear phase wavelet transforms in low bit-rate image coding. IEEE Transact. Image Process. 5 (5), 689–704.

Daubechies, I., 1992. Ten lectures on wavelets, SIAM Press, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61.

Johnsen, O., Shentov, O.V., Mitra, S.K., 1990. A technique for the efficient coding of the upper bands in subband coding of images, In: Proc. ICASSP. Albuquerque, NM, USA, pp. 2097–2100.

Kronander, T., 1989. Subband coding of monochrome and color images, Ph.D thesis, Delft University of Technology, Delft, The Netherlands.

Liu, C.-M., Wang, C.-N., 1998. Perceptual interband correlation for octave subband coding of images. In: Proc. ISCAS. vol. 4, Monterey, LA, USA, pp. 166–169.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11 (7), 674–693.

Mohsenian, N., Nasrabadi, N.M., 1994. Edge-based subband VQ techniques for images and video. IEEE Trans. Circuits Syst. Video Technol. 4 (1), 53–67.

MPEG Video Group, Information technology—Coding of audio-visual objects—Part 2: Visual & Amendment 1:Visual extension, International Standard, ISO/IEC JTC1/SC 29/WG 11 14496-2: 1999/ Amd. 1:2000(E), Jan. 2000.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1990. Numerical Recipes in C: The Art of Scientific Computing, second ed. Cambridge University Press, Cambridge.

Safranek, R.J., Johnston, J.D., 1989. A perceptually tuned subband image coder with image dependent quantization and post-quantization data compression. Proc. IEEE ICASSP 3, 1945–1948.

Shapiro, J.M., 1993. Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans. Signal Process. 41 (12), 3445–3462.

Vaidyanathan, P.P., Hoang, P.Q., 1988. Lattice structures for optimal design and robust implementation of two-channel perfect-reconstruction QMF banks. IEEE Trans. Acoustics, Speech, Signal Process. 36 (1), 81–94.

Vaisey, J., 1995. Subband prediction using leakage information in image coding. IEEE Trans. Commun. 43 (2/3/4), 216–221.

Wang, C.-N., Liu, C.-M., Tihao Chiang, 2001. Improved mpeg-4 visual texture coding using perceptual dithering for transparent image coding, In: Proc. of The Second IEEE Pacific-Rim Conference On Multimedia, Beijing, Oct. 24–26, pp. 134–141.

Woods, J.W., O'Neil, S.D., 1986. Subband coding of images. IEEE Trans. Acoustics, Speech, Signal Process. ASSP-34 (5), 1278–1288.

Chung-Neng Wang was born in PingTung, Taiwan, in 1972. He received the B.S. degree and Ph.D. degree in computer science and information engineering from National Chiao-Tung University (NCTU), HsinChu, Taiwan in 1994 and 2003, respectively. He joined the faculty at National Chiao-Tung University in Taiwan, ROC in January 2003. His current research interests are video/image compression, motion estimation, video transcoding, and streaming.

Chi-Min Liu received the B.S. degree in electrical engineering from Tatung Institute of Technology, Taiwan, ROC in 1985, and the M.S. degree and Ph.D. degree in electronics from National Chiao Tung University, Hsinchu, Taiwan, in 1987 and 1991, respectively.

He is currently a Professor of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include video/audio compression, speech recognition, radar processing, and application-specific VLSI architecture design.

Tihao Chiang was born in Cha-Yi, Taiwan, Republic of China, 1965. He received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1987, and the M.S. degree in electrical engineering from Columbia University in 1991. He received his Ph.D. degree in electrical engineering from Columbia University in 1995. In 1995, he joined David Sarnoff Research Center as a Member of Technical Staff. Later, he was promoted as a technology leader and a program manager at Sarnoff. While at Sarnoff, he led a team of researchers and developed an optimized MPEG-2 software encoder. For his work in the encoder and MPEG-4 areas, he received two Sarnoff achievement awards and three Sarnoff team awards.

Since 1992 he has actively participated in ISO's Moving Picture Experts Group (MPEG) digital video coding standardization process with particular focus on the scalability/compatibility issue. He is currently the co-editor of the part 7 on the MPEG-4 committee. He has made more than 50 contributions to the MPEG committee over the past 10 years. His main research interests are compatible/scalable video compression, stereoscopic video coding, and motion estimation. In September 1999, he joined the faculty at National Chiao-Tung University in Taiwan, ROC. Dr. Chiang is currently a senior member of IEEE and holder of 9 US patents and 26 European and worldwide patents. He was a co-recipient of the 2001 best paper award from the IEEE Transactions on Circuits and Systems for Video Technology. He published over 30 technical journal and conference papers in the field of video and signal processing.