https://doi.org/10.2352/ISSN.2470-1173.2018.14.HVEI-521

Optimum Space-Frequency Partition in Subband Image Coding with Human Visual Sensitivity and Region-of-Interest

Haruhiko Miyazaki and Masashi Kameda; Iwate Prefectural University, Graduate School of Software and Information Science; Takizawa, Iwate, Japan

Abstract

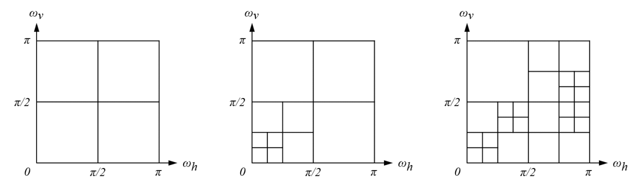
Subband coding is a powerful means for highly efficient image compression. In order to improve the coding performance of subband image coding, we recently have proposed the optimum space-frequency partition coder (OSFP) that optimizes the following three factors in the rate-distortion sense: the frequency band partition with a small number of subbands, quantization and the spatial segmentation to exclude redundant pixels. However, an encoded image obtained by OSFP is not necessarily optimal in subjective image quality because the three factors are optimized to minimize the mean square error (MSE). In this paper, we present a new OSFP that obtains a high quality coded image subjectively by optimizing the three factors so that MSE weighted by considering both the human visual sensitivity and a region-of-interest of human is minimized. Experimental results show that the quality of encoded images obtained by the proposed method has higher subjectively than them of both the conventional OSFP and JPEG2000 by the mean opinion score (MOS).

Introduction

Subband image coding [1] is a powerful method for highly efficient image compression without suffering from occurrences of blocking artifacts by using the discrete cosine transform (DCT) such as JPEG coder. In the basic scheme of subband image coding, at first, the 2-dimensional frequency domain of an input image is divided into four subbands (**Figure 1(a)**) by applying the analysis filter bank. Secondly, the quantization is applied to each subband signal. Finally, an entropy coder such as an arithmetic coder [2] is designed in each subband considering the distribution of subband signal independently, thus all coefficients in each subband are encoded into the bit-stream. Here, in order to improve the coding performance of subband image coding, it is desirable to divide the 2-dimensional frequency domain adaptively considering the characteristics of an input image [1].

The discrete wavelet transform (DWT) [3] has been adopted in JPEG2000 coder [4] that is an extension of original subband image coding and can improve the coding performance by partitioning the lowest frequency subband (called LL subband) recursively (**Figure 1(b)**) because the power of a typical real-world image is concentrated in lower frequency domain. However, for any images including more high frequency components (e.g., stripe, texture), DWT cannot improve sufficiently the coding performance due to the recursive partitioning of LL subband only.

Wavelet Packet (WP) [5] offers to adapt the partitioning of the 2-dimensional frequency domain according to the characteristics of an input image by allowing to further partition for all subbands and not just LL subband (**Figure 1(c)**). It is important to choose the best partition pattern from among enormous possible patterns, therefore the application of image compression based on WP employs the technique called to “best basis algorithm” to search the best partition pattern. At first, K. Ramchandran *et al*. demonstrated the employment of rate-distortion optimization criteria of WP for classical wavelet-based image compression schemes [5]. Specifically, it can determine both the optimum partition pattern and the optimum combination of quantizers for each subband in terms of the rate-distortion sense. So far, today’s various image compression methods based on WP (e.g., [6]-[8]) have been introduced incorporating the principle of [5]. However, they are a system with a higher computational complexity than DWT. This is because of increasing the computational costs required for not only determination of the optimum partition pattern but also the design of both quantizer and entropy coder for the large number of subbands, which compose the partition pattern. Therefore, in order to improve the coding performance of subband image coding under the condition of lower computational costs, it is desirable to obtain an appropriate partition pattern considering the characteristics of an input image with a smaller number of subbands.



(a) Original (b) DWT (c) WP

Figure 1. *Examples of the partition pattern on the 2-dimensional frequency domain by original subband image coding, discrete wavelet transform (DWT) and Wavelet Packet (WP). Subbands are each region divided by block.*

In our previous study [9], we have proposed the optimum frequency band partition coder (OFBP) that determines an adaptive partition pattern on the 2-dimensional frequency domain with a given small number of subbands to maximize the coding gain considering the variances of an input image. In order to improve OFBP, we have presented OFBP-RD [10] that determines both the optimum partition pattern on the 2-dimensional frequency domain and the optimum combination of scalar quantizers for each subband in terms of the rate-distortion sense, i.e., minimize the total quantization distortion under a desired total bit rate constraint. Furthermore, we have developed the optimum space-frequency partition coder (OSFP) [11] that optimizes not only the frequency band partition and quantization but also segmentation on the 2-dimensional spatial domain of each subband, which excludes redundant coefficient pixels in terms of the rate-distortion sense. The experimental results show that the coding performance of OSFP is significantly superior to them of OFBP [9], OFBP-RD [10], DWT and WP [5], despite the number of subbands in partition pattern by OSFP is much smaller than DWT and WP. However, since the optimization problem of OSFP is equivalent to the

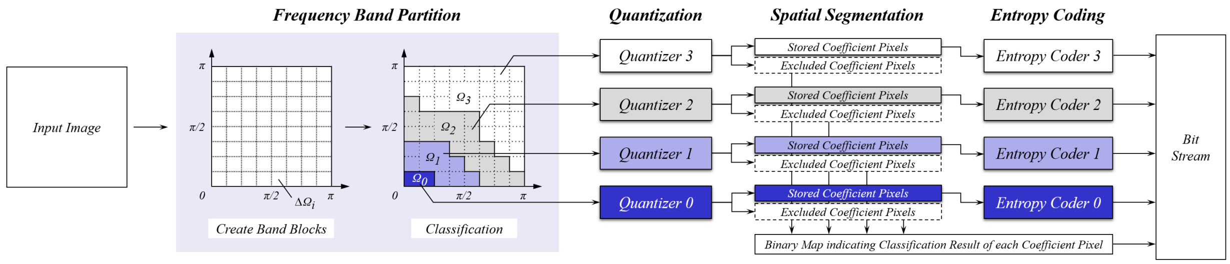


Figure 2. *Configuration of the optimum space-frequency partition coder (OSFP) under the condition of $N = 64$ and $M = 4$.*

minimization problem of the mean square error (MSE) between input image and encoded image, OSFP is not optimal in terms of subjective image quality. This reason being that the MSE is not always well correlated with image quality assessment of the human visual system.

In this paper, we present a new OSFP coder to obtain a high quality coded image subjectively. In our method, after the RGB components of the color image are transformed to the YCrCb components, a new distortion function which is defined as a weighted MSE by considering both the human visual sensitivity and a region-of-interest (ROI) of human is minimized on each color component domain independently under a given rate constraint. Here, a weighted function is derived according to both the spatial frequency characteristic also known as the modulation transfer function (MTF) [12] and a saliency map to estimate ROI of a human [13]. Moreover, ROI and the other region (non-ROI) on the spatial domain in each subband can be quantized separately by the optimum quantizer considering the weighted distortion function. Finally, experimental results show that encoded images obtained by the proposed OSFP coder are superior to them by our conventional OSFP and JPEG2000 at the same bit rate in subjective image quality by the mean opinion score (MOS).

Our Previous Works

The original OFBP [9] divides the 2-dimensional frequency domain of an input image into M subbands to maximize the coding gain, where the number of subbands M is a small number given by users. When OFBP is implemented to a real-world image, at first, the 2-dimensional frequency domain is divided into a set of N small square regions with the same bandwidth by applying a filter bank such as the QMF [1]. These small frequency regions are called to the band blocks $\Delta\Omega_i (i = 0, 1, \cdots, N-1; N \gg M)$. Secondly, N band blocks are classified into M groups which correspond to subband $\Omega_k (k = 0, 1, \cdots, M-1)$ to maximize the coding gain. Finally, the quantization and entropy coding are applied to each subband Ω_k independently. However, the partition pattern obtained by OFBP is not optimal in terms of the rate-distortion sense for image compression. The reason why is that the partition pattern remains fixed despite each band block signal is changing through quantization for variable bit rate. In other words, the partition pattern cannot be adapted to changing of bit rate.

In order to solve the above problem, we have presented the OFBP-RD [10] that determines simultaneously both the optimum band partition pattern on the 2-dimensional frequency domain and the optimum combination of quantizers for each subband so that the total quantization distortion is minimized when an arbitrary bit rate budget is given. The experimental results show that the coding performance of OFBP-RD is improved from the original OFBP [9] and DWT, however, inferior to WP [5].

To improve the coding performance of OFBP-RD while not increasing the number of subbands, we have developed the OSFP [11] that determines simultaneously not only the optimum frequency band partition pattern and the optimum combination of quantizers but also the optimum segmentation pattern on the 2-dimensional spatial domain of each subband in terms of the rate-distortion sense. The segmentation on the 2-dimensional spatial domain corresponds to classify coefficient pixels in each subband into “a set of stored pixels” and “a set of excluded pixels”. Then, only the former set is sent to the decoder side and coefficient pixels in the latter set is interpolated by the zero value on the decoder side, where a binary map indicating classification result of each coefficient pixel needs to send to the decoder side as an additional information. **Figure 2** illustrates the configuration of OSFP coder under the condition of $N = 64$ and $M = 4$.

Let I_k, q_k and S_i denote a set of band blocks in subband Ω_k (i.e., the frequency band partition pattern), quantizer for subband Ω_k (i.e., the combination of quantizers) and a set of stored pixels in band block $\Delta\Omega_i$ (i.e., the spatial segmentation pattern), respectively. They are optimized by solving the following three equation iteratively in the rate-distortion sense:

$$I_k = arg \min_{I_k} \left[\sum_{k=0}^{M-1} \sum_{i \in I_k} \sum_{x} \Delta N_i \Delta\Omega_i(x) \log_2 \frac{\Delta\Omega_i(x)}{\Omega_k(x)} \right], \tag{1}$$

$$q_k = arg \min_{q_k} \left[\begin{array}{c} \displaystyle\sum_{i \in I_k, j \in S_i} \left(x_{i,j} - \hat{x}_{i,j}^{q_k} \right)^2 \\ +\lambda \left\{ -N_k \displaystyle\sum_{x} \Omega_k(x) \log_2 \Omega_k(x) \right\} \end{array} \right], \tag{2}$$

and $S_i \leftarrow j$-th pixel in band block $\Delta\Omega_i$ if

$$x_{i,j}^2 \leq \left[\left(x_{i,j} - \hat{x}_{i,j}^{q_k} \right)^2 + \lambda \left\{ -\log_2 \Omega_k(\hat{x}_{i,j}^{q_k}) \right\} \right]. \tag{3}$$

In Equation (1), (2) and (3), $\Delta\Omega_i(x)$ and $\Omega_k(x)$ is probability of coefficient value x in band block $\Delta\Omega_i$ and subband Ω_k, respectively. And $x_{i,j}$ and $\hat{x}_{i,j}^{q_k}$ is unquantized and quantized (by quantizer q_k) coefficient value of j-th pixel in band block $\Delta\Omega_i$, respectively. Also, ΔN_i and N_k is the number of stored pixels in band block $\Delta\Omega_i$ and subband Ω_k, respectively, and λ is the Lagrange multiplier. The experimental results show that the coding performance of OSFP is about 2.0 [dB] larger than OFBP-RD [10] and WP [5] despite a small number of subbands such as $M = 5$.

However, encoded images obtained by OSFP are not necessarily appropriate in terms of human subjective image quality because the quantization distortion measure in the above equations is defined as $(x_{i,j} - \hat{x}_{i,j}^{q_k})^2$, i.e., MSE between unquantized and quantized coefficient values. MSE is known as distortion measure which is not always justified when compared to the image quality assessment of the human. Therefore, it is necessary to determine three factors I_k, q_k and S_i by using a new distortion measure to improve the image quality of encoded image subjectively.

Optimum Space-Frequency Partition with Human Visual Sensitivity and Region-of-Interest

We propose a new OSFP coder to obtain the encoded image with high image quality subjectively. At first, the proposed method transforms the RGB components of an input color image to the YCrCb components, then determines three factors I_k, q_k and S_i on each color component domain separately so that a weighted MSE as an alternative distortion measure is minimized subject to constraint on a given bit rate. Here, we define a weighted function for MSE by considering both the spatial frequency characteristic of the human visual sensitivity well known as the modulation transfer function (MTF) [12] and a saliency map to estimate a region-of-interest (ROI) of when a human observes a given image [13]. Moreover, each subband is divided into ROI and non-region-of-interest (non-ROI) on the spatial domain by reflecting a result of the saliency map, then the ROI and the non-ROI in each subband are quantized separately by the optimum quantizers considering the weighted MSE.

The RGB color components of an input color image are transformed to the YCrCb color components by

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.500 & -0.419 & 0.081 \\ -0.169 & -0.331 & 0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{4}$$

When an input color image is given, a set of three factors I_k, q_k and S_i will be determined independently on each color component Y, Cr and Cb domain.

A. Modulation Transfer Function

The spatial frequency characteristic of the human visual sensitivity is well known as the modulation transfer function (MTF) [12]. MTF is a characteristic of any structure that is periodic across position on spatial domain, and is a measure of how often sinusoidal components of the structure repeat per one degree of visual angle (called cycles per degree [cpd]). Although MTF has a bandpass characteristic, its details vary depending on the measurement conditions. We use a MTF based on a retinal neuro circuit model [12], and then the MTF can be represented by

$$S(f) = 1.5exp(-\sigma^2\omega^2/2) - exp(-2\sigma^2\omega^2) \tag{5}$$

, where $\sigma = 2$, $\omega = 2\pi f/60$, $f = \sqrt{u^2+v^2}/2\pi$, and u and v denote the horizontal and vertical spatial frequencies, respectively. Also, f means the cycles per one degree of visual angle, i.e., [cpd]. Furthermore, we will consider the anisotropy of the 2-dimensional MTF whose this characteristic can be approximately represented by

$$O(\theta) = \begin{cases} 1.0 & : f < f_p \\ 0.5(1+cos^4 2\theta) & : f \geq f_p \end{cases} \tag{6}$$

, where $\theta = tan^{-1}(v/u)$, and $f_p = 3.86$ [cpd] which is the cycles per degree at which $S(f)$ becomes maximum. Therefore, the MTF considering the anisotropy on the 2-dimensional frequency domain is defined by the product of Equation (5) and (6), i.e.,

$$\psi(u,v) = S(f)\cdot O(\theta)\ . \tag{7}$$

In order to define Equation (7) in our experiments, pixels per degree [ppd] need to be calculated according to the measurement condition such as a display resolution and a viewing distance. Then, u and v are normalized so that the cycles per degree of the minimum period (i.e., 2 pixels) occurs when these two parameters are π. Here, pixels per degree [ppd] can be calculated by

(a) "kodim04" (b) Saliency map (c) Segmentation

Figure 3. *Original test image "kodim04"(512×768), its saliency map and result of segmentation of ROI and non-ROI in band block (the actual size is 64×96) when the region ratio of ROI is given as $\rho = 1/5$.*

$$ppd = 1/tan^{-1}\left(\frac{2.54\cdot D}{\sqrt{H^2+V^2}\cdot d}\right) \tag{8}$$

, where D is the diagonal size (inch) of the display, and H and V are the number of pixels in the horizontal and vertical direction, respectively. And d is the viewing distance (cm) from the eyes to the screen.

We will employ the MTF on the 2-dimensional frequency domain obtained by Equation (7) to define the weighted function. Specifically, when an input image with a pixel length $(X\times Y)$ is given, we define a weighted function corresponding to each band block $\Delta\Omega_i$ for reflecting the MTF,

$$\bar{\psi}_i = \frac{N}{XY}\sum_{(\pi m/X,\pi n/Y)\in\Delta\Omega_i}\psi(\pi m/X, \pi n/Y) \tag{9}$$

, where $(0\leq m<X, 0\leq n<Y)$. The quality of sensitive frequency bands for the human visual system can be kept preferentially by weighting $\bar{\psi}_i$ into a quantization distortion function of each band block $\Delta\Omega_i$.

B. Region-of-Interest based on Saliency Map

The region-of-interest (ROI) is a subset area in an image identified for a particular purpose (e.g., image compression, object recognition). A saliency map is a topographically arranged map that represents visual saliency of a corresponding visual scene, and its saliency model is calculated from some image features [13]. Therefore, the saliency map is possible to represent quantitatively ROI when a human observes a given image. **Figure 3(a)** shows the image "kodim04"(512×768) in the Kodak standard image dataset in the CIPR still images [14] and **Figure 3(b)** shows the obtained saliency map for this image by the saliency detection algorithm in [13]. In **Figure 3(b)**, regions with higher luminance value correspond to ROI.

We will also employ the saliency map on the 2-dimensional spatial domain to define the weighted function. Specifically, a saliency map $\phi(X\times Y)$ for an input image with a pixel length $(X\times Y)$ is generated by the algorithm in [13], then the $\phi(X\times Y)$ is resized to the same size of band block by applying the pixel average method. The resized $\phi(X\times Y)$ denotes $\bar{\phi}_j (j=0,1,\cdots,XY/N)$, where the position of j-th pixel on $\bar{\phi}_j$ corresponds to the position of j-th pixel on band block. A coefficient value of the $\bar{\phi}_j$ is weighted into a quantization distortion function of j-th pixel in all band blocks for reflecting the saliency map. Next, in order to quantize both ROI and non-region-of-interest (non-ROI) in each subband, the proposed method divides $\bar{\phi}_j$ into two areas of ROI and non-ROI definitely. The goal of this quantization strategy is to keep preferentially the image quality in ROI by applying a quantizer with fine quantization level to ROI instead of applying a quantizer with coarse quantization to non-ROI. This goal is achieved by both the incorporation of the weighted function based on $\bar{\phi}_j$ and the region segmentation of ROI and non-ROI. In the proposed method, we divide the region of $\bar{\phi}_j$ into ROI and non-ROI by the following processing:

- Step 1: Set $\alpha_j \leftarrow 0$ $(j = 0, 1, \cdots, XY/N)$ and $x \leftarrow \lfloor x_{max} \rfloor$, where α_j is a binary map indicating ROI $[\alpha_j = 1]$ or non-ROI $[\alpha_j = 0]$. Also, x_{max} is the maximum coefficient value of $\bar{\phi}_j$ and $\lfloor * \rfloor$ is the floor function.
- Step 2: $\alpha_j \leftarrow 1$ $(\forall j \in K)$, where K is a set of number j of when $x = \lfloor \bar{\phi}_j \rfloor$.
- Step 3: Set $x \leftarrow x - 1$ and go back to Step 2 if ratio of pixels with $\alpha_j = 1$ less than ρ, else go to Step 4. ρ is the region ratio of ROI to the whole region of $\bar{\phi}_j$, is given by users.
- Step 4: Dilate the area of ROI (i.e., $\alpha_j = 1$) by the dilation operation with the structuring element of matrix (3×3). Then, α_j is defined as a result of region segmentation of ROI and non-ROI.

ROI and non-ROI in each band block is defined as a set of j-th pixels when $\alpha_j = 1$ and $\alpha_j = 0$, respectively. A determination of quantizer for ROI and non-ROI in each subband will be described in the following subsection.

C. Determination of Optimum Factors based on Weighted Function

We define the weighted function for the quantization distortion measure of j-th pixel in band block $\Delta\Omega_i$ as

$$W_{i,j} = \| \bar{\psi}_i \| \cdot \| \bar{\phi}_j \| \tag{10}$$

, where $\| * \|$ is the function to normalize the value from 0 to 1. The coefficient value of $W_{i,j}$ becomes higher when a salient pixel in band block corresponding to the MTF with a high value is specified. Also, $W_{i,j}$ becomes lower when j-th pixel is an unremarkable pixel such as belonging to non-ROI even if band block $\Delta\Omega_i$ corresponds to the MTF with a high value. Similarly, $W_{i,j}$ becomes lower when band block $\Delta\Omega_i$ corresponds to the MTF with a low value even if j-th pixel is a salient pixel. The proposed method determines three factors I_k, q_k and S_i for each color component Y, Cr and Cb domain independently so that the weighted MSE by $W_{i,j}$ is minimized subject to constraint on a given bit rate budget. Here, regarding the combination of quantizers, the optimum quantizers for ROI and non-ROI in each subband are determined in order to keep the quality of ROI preferentially instead of sacrificing the quality of non-ROI. However, in order to not increase the computational costs, it is desirable that the entropy coder is not designed independently for ROI and non-ROI, but is designed in each subband as with our conventional methods. Therefore, it is necessary to select the optimum quantizers for ROI and non-ROI in each subband while considering the trade-off relationship between total rate and total distortion (i.e., the weighted MSE by $W_{i,j}$) in the subband.

Let $q_{k,1}$ and $q_{k,0}$ denotes quantizer for ROI and non-ROI in subband Ω_k, respectively. The optimum combination of $q_{k,1}$ and $q_{k,0}$ are determined by

$$(q_{k,1}, q_{k,0}) = arg \min_{q_{k,1}, q_{k,0}} \left[\begin{array}{c} \sum_{i \in I_k, j \in S_i} W_{i,j}^2 \left(x_{i,j} - \hat{x}_{i,j}^{q_{k,\alpha_j}} \right)^2 \\ + \lambda \left\{ -N_k \sum_x p_k(x) \log_2 p_k(x) \right\} \end{array} \right] \tag{11}$$

, where α_j is the binary map indicating that j-th pixel in band block is a pixel in ROI $[\alpha_j = 1]$ or non-ROI $[\alpha_j = 0]$, obtained in the previous subsection. Also, $\hat{w}_{i,j}^{q_{k,\alpha_j}}$ represents the coefficient value quantized by $q_{k,1}$ and $q_{k,0}$ when j-th pixel belongs to ROI and non-ROI, respectively. In Equation (11),

$$p_k(x) = \frac{1}{N_k} \sum_{i \in I_k} \Delta N_i \Delta p_i(x) \tag{12}$$

, where $\Delta p_i(x)$ is a special mixture probability distribution of both ROI and non-ROI in band block $\Delta\Omega_i$, calculated by

$$\Delta p_i(x) = \frac{\Delta N_{i,1} \cdot \Delta\Omega_{i,1}^{q_{k,1}}(x \cdot \mid q_{k,1} \mid) + \Delta N_{i,0} \cdot \Delta\Omega_{i,0}^{q_{k,0}}(x \cdot \mid q_{k,0} \mid)}{\Delta N_i}. \tag{13}$$

In Equation (13), $\Delta\Omega_{i,1}^{q_{k,1}}(-)$ represents the probability distribution of ROI and non-ROI which is quantized by quantizer $q_{k,1}$ and $q_{k,0}$ in band block $\Delta\Omega_i$, respectively. $\Delta N_{i,1}$ and $\Delta N_{i,0}$ is the number of stored pixels in ROI and non-ROI in band block $\Delta\Omega_i$, respectively (i.e., ΔN_i is equivalent to the sum of $\Delta N_{i,1}$ and $\Delta N_{i,0}$). Also, $\mid q_{k,0} \mid$ and $\mid q_{k,1} \mid$ is the quantization step-size of scalar quantizer $q_{k,0}$ and $q_{k,1}$, respectively. $\Delta p_i(x)$ is generated by integrating $\Delta\Omega_{i,1}^{q_{k,1}}(-)$ and $\Delta\Omega_{i,0}^{q_{k,0}}(-)$ reconstructed so that histogram bins of two probability distributions are concentrated around the zero value because if two probability distributions quantized by a different quantizer are integrated simply, the entropy of $\Delta p_i(x)$ (i.e., rate of band block $\Delta\Omega_i$) will be increased significantly. The entropy coder in subband Ω_k is designed according to the probability distribution $p_k(x)$, which is also integrated $\Delta p_i(x) (i \in I_k)$. In the decoder side, each decoded coefficient value $y_{i,j}$ of j-th pixel in band block $\Delta\Omega_i$ is reconstructed by

$$y_{i,j} = \mid q_{k,\alpha_j} \mid \cdot \hat{y}_{i,j} \tag{14}$$

, where $\hat{y}_{i,j}$ is a coefficient value decoded by an entropy decoder directly. For instance, when $\hat{y}_{i,j} = 2$ is given under the condition of $q_{k,1} = 2$ and $q_{k,0} = -5$, if j-th pixel belongs to ROI $(\alpha_j = 1)$ and non-ROI $(\alpha_j = 0)$, $y_{i,j} = 4$ and $y_{i,j} = -10$ is reconstructed, respectively. Moreover, other factors I_k and S_i are also determined by using $W_{i,j}$, $p_k(x)$ and $\Delta p_i(x)$ as the following:

$$I_k = arg \min_{I_k} \left[\sum_{k=0}^{M-1} \sum_{i \in I_k} \sum_x \Delta N_i \Delta p_i(x) \log_2 \frac{\Delta p_i(x)}{p_k(x)} \right], \tag{15}$$

and $S_i \leftarrow$ j-th pixel in band block $\Delta\Omega_i$ if

$$W_{i,j}^2 \cdot x_{i,j}^2 \leq \left[W_{i,j}^2 \left(x_{i,j} - \hat{w}_{i,j}^{q_{k,\alpha_j}} \right)^2 + \lambda \left\{ -\log_2 p_k \left(\hat{x}_{i,j}^{q_{k,\alpha_j}} \right) \right\} \right]. \tag{16}$$

In the proposed method, the factors I_k, $q_{k,1}$, $q_{k,0}$ and S_i are determined by solving Equation (15), (11) and (16) iteratively.

In the proposed method, I_k is determined by the classification of band blocks using the special mixture probability distribution $\Delta p_i(x)$ and $p_k(x)$. Also, S_i is determined by comparing between the unquantized coefficient value $x_{i,j}$ weighted by $W_{i,j}$ and the Lagrangian cost function based on both weighted MSE by $W_{i,j}$ and $p_k(x)$. Quantizers $p_{k,1}$ and $p_{k,0}$ are determined separately for ROI and non-ROI so that the Lagrangian cost function mentioned above is minimized. Also, α_j indicating classification result of ROI and non-ROI is encoded using run-length encoding and thus sent to the decoder side as the additional information.

Experimental Results and Discussion

We apply the proposed method, our conventional method and JPEG2000 [4] to three test images "kodim04"(512×768) [14], resized "p22"(640×800) in the image dataset provided by the AIC JPEG ad-hoc group [15] and "girl"(480×512) in the CIPR still images [14]. These test images are all portrait shown at the top-left in **Figure 5**, and they are easy to distinguish ROI when participants observe each test image. In the proposed method and our conventional method in common, the following conditions are given: $N = 64$, $M = 5$ and the quantization step-size q of scalar quantizer takes values from the set $(q : q = 1, 2, \cdots, 32)$. Also, parameters D, H, V and d for the MTF are shown in **Table 2**, and the region ratio of ROI in the proposed method uses $\rho = 1/5$ considering the contents of each image. For example, **Figure 3(c)** shows the result of region segmentation of ROI and non-ROI in band block for the image "kodim04", where white and black regions represent ROI and non-ROI, respectively. Also, the decomposition level of DWT in JPEG2000 is set three as with the proposed method and our conventional method.

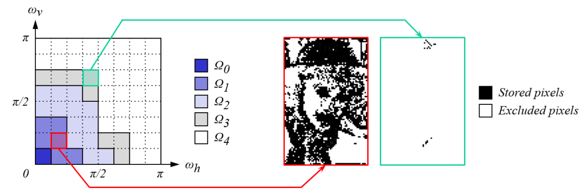
A. Experimental Results

We have shown the results of three factors I_k, $q_{k,1}$, $q_{k,0}$ (or q_k only) and S_i on color component Y domain, which is concentrated the energy of the color image. **Figure 4** illustrates the results of the 2-dimensional frequency band partition pattern by (a) the proposed method and (b) our conventional method for the image "kodim04" when 0.10 [bit/pel] is given as the desired bit rate budget, respectively. Also, **Figure 4** shows the results of the 2-dimensional spatial segmentation patterns in two band blocks with lower and higher frequency components, where black and white pixels represent stored pixels and excluded pixels, respectively. **Table 1** shows the results of the combination of the quantization step-size of scalar quantizers applied to (a) ROI and non-ROI in each subband in **Figure 4(a)** and (b) each subband in **Figure 4(b)**. In **Table 1**, subband Ω_4 is not quantized in both methods because all pixels in subband are excluded.

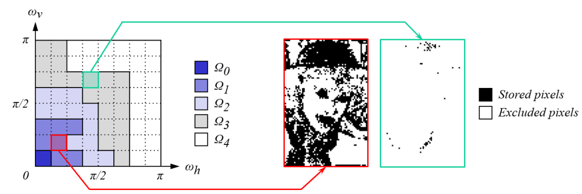
From the results in **Figure 4**, both the frequency band partition pattern and the spatial segmentation pattern by the proposed method are different from them by our conventional method. Especially, when comparing the two spatial segmentation patterns in the two band blocks shown by red frame, it is observed that pixels belonging to ROI (i.e., face region) in band block in (a) are stored more than (b). The reason why is that the distortion of pixels belonging to ROI is weighted strongly by the effect of $\bar{\phi}_j$ in Equation (10). On the other hand, pixels in band block shown by green frame in (a) are excluded more than (b) because the band block shown by green frame corresponds to the MTF with a low value. Next, from the results in **Table 1(a)**, ROIs in subband Ω_0, Ω_1 and Ω_2 are quantized with fine quantization level compared with one for non-ROI. And, both ROI and non-ROI in subband Ω_3 corresponding to the MTF with a low value are quantized with the coarsest quantization level. Consequently, it is considered that the quality of ROIs in subband Ω_0, Ω_1 and Ω_2 can be kept instead of sacrificing the quality of non-ROIs in the same subband and subband Ω_3 by incorporation of the weighted function $W_{i,j}$.

B. Evaluation of Image Quality

We carried out experiments on subjective evaluation of image quality, and compared performances by the proposed method with those by our conventional method and JPEG2000. **Table 2** shows the conditions of the experiments on subjective evaluation of image quality. Each image was evaluated on a five-point scale shown in the lower part of **Table 2**. The mean opinion score (MOS) is



(a) The proposed method



(b) Our conventional method

Figure 4. *The results of the 2-dimensional frequency band partition patterns and examples of the results of the 2-dimensional spatial segmentation patterns in two band blocks on color component Y domain for the image "kodim04" (0.10 [bit/pel]).*

Table 1. The combination of the quantization step-size of scalar quantizers for each subband shown in Figure 4.

Band Partition Pattern		Ω_0	Ω_1	Ω_2	Ω_3	Ω_4
(a)	ROI	4	3	5	32	–
	non-ROI	5	5	7	32	–
(b)		5	5	6	5	–

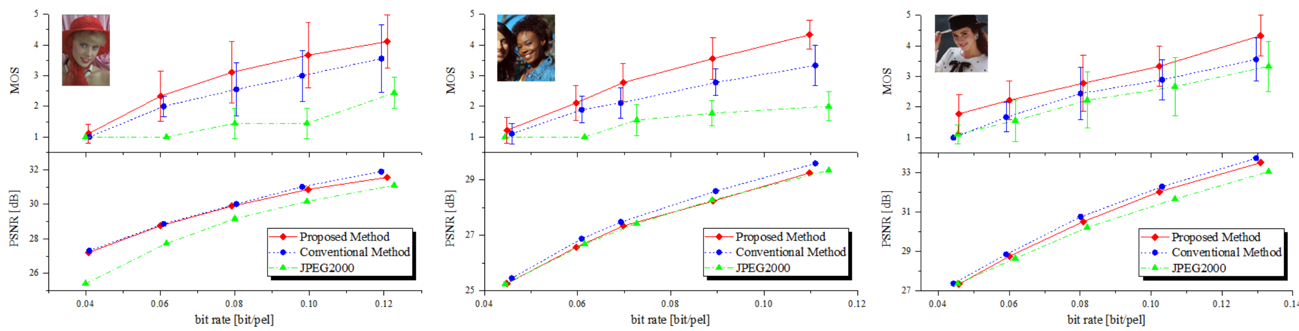
given by

$$MOS = \frac{1}{T} \sum_{t=0}^{T-1} a_t \ . \tag{17}$$

, where T is the number of participants, and a_t is the t-th subject's score. **Figure 5** shows the results of comparisons of the MOS versus bit rate characteristics by the proposed method, our conventional method and JPEG2000. In order to calculate MOS values for different bit rates, we chose five coded images with each different bit rate as a set of images displayed to participants. **Figure 5** also shows the results of comparisons of the PSNR versus bit rate characteristics as the objective evaluation of image quality. From the results in **Figure 5**, the proposed method has higher MOS value than both our conventional method and JPEG2000 at the same bit rate although the PSNR values are nearly the same or lower than our conventional method. Consequently, it is obvious that the proposed method is better than our conventional method and JPEG2000 in terms of subjective image quality.

Conclusion

In this paper, we have presented the optimum space-frequency partition coder to obtain the encoded image with high image quality subjectively. Our approach is to determine the optimum three factors: the optimum partition pattern on the 2-dimensional frequency domain, the optimum combination of quantizers and the optimum segmentation pattern on the 2-dimensional spatial domain so that total quantization distortion weighted by considering both the human visual sensitivity and a region-of-interest (ROI) of human is minimized subject to a desired total bit rate constraint. We first have defined the weighted



(a) “kodim04” (b) “p22” (c) “girl”

Figure 5. *Comparisons of the MOS and PSNR vs. bit rate characteristics by the proposed method, our conventional method and JPEG2000.*

Table 2. Condition of experiments for subjective evaluation of image quality.

Evaluation method		Absolute category rating (ACR)
Display resolution	D (inch)	27 (inch)
	$H \times V$ (pixels)	1920 $\times$ 1080 (pixels)
Viewing distance d (cm)		50 (cm)
Display time per one image		10 (sec)
Number of participants T		10

Evaluation scale	Evaluation words
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

function according to both the spatial frequency characteristic of the human visual sensitivity and the saliency map, then we have proposed the method to determine separately the optimum quantizers for ROI and non-region-of-interest (non-ROI) in each subband so that the weighted quantization distortion is minimized. Finally, we applied the proposed method, our conventional method and JPEG2000 to three real-world test images. Then, we carried out experiments on subjective evaluation of image quality of encoded images, and compared the mean opinion score (MOS) versus bit rate characteristics of the proposed method with them of the others. Experimental results show that encode images obtained by the proposed method has higher MOS value than them of our conventional method and JPEG2000.

References

[1] J.W. Woods and S.D. O’Neil, “Subband coding of images,” IEEE Trans. on Acoustics Speech and Signal Processing, vol.34, no.5, pp.1278-1288, 1986.

[2] I. Witten, R. Neal and J. Cleary, “Arithmetic coding for data compression,” Commun. ACM, vol.30, pp.520-540, 1987.

[3] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, “Image coding using wavelet transform,” IEEE Trans. on Image Processing, vol.1, no.2, pp.205-220, 1992.

[4] A. Skodras, C. Christopoulos and T. Ebrahimi, “The JPEG 2000 still image compression standard,” IEEE Signal Processing Magazine, vol.18, no.5, pp.36-58, 2001.

[5] K. Ramchandran and M. Vetterli, “Best wavelet packet basis in a rate-distortion sense,” IEEE Trans. on Image Processing, vol.2, no.2, pp.160-175, 1993.

[6] N.M. Rajpoot, R.G. Wilson, F.G. Meyer and R.R. Coifman, “Adaptive wavelet packet basis selection for zerotree image coding,” IEEE Trans. on Image Processing, vol.12, no.12, 1460-1472, 2003.

[7] T. Stütz, B. Mühlbacher and A. Uhl, “Efficient and rate-distortion optimal wavelet packet basis selection in JPEG2000,” IEEE Trans. on Multimedia, vol.14, no.2, pp.264-277, 2012.

[8] F. Wenchao and X. Wang, “An image compression method based on orthogonal wavelet packet transform,” Boletín Técnico, vol.55, no.1, pp.18-25, 2017.

[9] M. Kameda, K. Ohtake and M. Miyahara, “2-dimensional optimum frequency band partition based on band blocks for subband image coding,” IEICE Trans., vol.J81-D-II, no.6, pp.1095-1107, 1998.

[10] H. Miyazaki and M. Kameda, “The optimum frequency band partition for subband image coding in the rate-distortion sense,” IEICE Trans., vol.J100-A, no.5, pp.205-217, 2017.

[11] H. Miyazaki and M. Kameda, “The optimum space-frequency partition for subband image coding in a rate-distortion sense,” Proc. of Picture Coding Symposium (PCS) 2016, 2016.

[12] M. Miyahara, H. Otani, H. Horita and T. Fujimoto, “Objective evaluation scale for image qualities (PQS); Local features and general-purpose applications,” IEICE Trans., vol.J73-B-I, pp.208-218, 1990.

[13] Q. Yan, L. Xu, J. Shi and J. Jia, “Hierarchical Saliency Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1155-1162, 2013.

[14] “CIPR still images,” http://www.cipr.rpi.edu/resource/stills/, Sep. 2017 access.

[15] “JPEG core experiment for the evaluation of JPEG XR image coding,” http://mmspg.epfl.ch/iqa, Sep. 2017 access.

Author Biography

Haruhiko Miyazaki graduated from the Faculty of Software and Information Science of Iwate Prefectural University in 2013 and completed the master’s program in 2015. He is now studying at the doctoral program. He has engaged in research on image coding.

Masashi Kameda graduated from the Electronics Department of Nagaoka University of Technology in 1990 and completed the M.E. program in 1992. He then became a research associate at Kobe University. He graduated from the doctoral program at Japan Advance Institute of Science and Technology in 1998, obtaining a D. Information Science degree. He is now a professor at Iwate Prefectural University. He had engaged in research on image compression and image processing.