OPTIMAL JPEG2000 ENCODER MECHANISM FOR LOW DELAY AND EFFICIENT DISTRIBUTION OF HDTV PROGRAMS

Sei Naito *Atsushi Koike* *Shuichi Matsumoto*

KDDI R&D Laboratories Inc.
2-1-15 Ohara, Kamifukuoka-shi
SAITAMA 356-8502 JAPAN

ABSTRACT

Motion JPEG2000 has been developed as a new video coding standard for motion pictures utilizing the still image coding standard JPEG2000. Motion JPEG2000 specifies a normative bitstream syntax that the decoder must recognize and allows flexible selection of detailed coding parameters at the encoder. The compression performance largely depends, therefore, on the encoder design. However, the simple implementation of JPEG2000 encoder might cause picture quality degradation due specifically to wavelet image coding of interlaced TV signals and unsatisfactory bit assignments. That degradation might then become conspicuous in subjective picture quality especially during low bit-rate coding. To overcome this problem, we investigated an optimal design for a Motion JPEG2000 encoder that can be used for primary distribution applications of HDTV programs, assuming that the available bit-rate is approximately 50Mbps.

Our study introduces advanced key technologies not yet recognized officially. Coding experiments using those technologies confirmed that a significant coding gain was achieved versus conventional encoding schemes.

1. INTRODUCTION

A new video coding standard for motion pictures utilizing JPEG2000 technology is specified as Motion JPEG2000 [1]. The advantage of employing JPEG2000 intraframe coding technology for encoding motion pictures is that the codec delay can be kept much smaller compared to conventional MC+DCT coding while maintaining highly efficient coding performance. From this viewpoint, the JPEG2000 encoder is expected to greatly improve primary distribution applications of TV signals where extremely high picture quality and very low delay transmission are required. However, simply design of the JPEG2000 encoder might cause picture quality degradation due specifically to wavelet image coding of interlaced TV signals and unsatisfactory bit assignments.

In this paper, 1080i HDTV is assumed as the typical video application and an optimal encoder design to maintain satisfactory picture quality at a bit-rate around 50Mbps are proposed. This proposed coding scheme is comprised of two key technologies an optimal rate control and adaptive coding type selection. Both of these new technologies can be incorporated into HDTV coding based on JPEG2000 while maintaining bitstream conformance. In the following, we describe proposed key technologies and evaluate total coding performance from results of coding experiments.

2. ASSUMED CODING STRUCTURE

The block diagram of the target encoder is shown in Figure 2.1. As a video input, 1080i HDTV signal is assumed, and JPEG2000 coding is applied for each video field or video frame sequentially. In this paper, the video unit for which JPEG2000 coding is applied is simply called picture. In general encoding schemes for motion pictures, a buffer model is specified for verification if the encoder output accommodates transmissions at a constant bit-rate (CBR) or not. This buffer is simply called a verification buffer in the following. One problem associated with introducing a buffer model is the increase in codec delay in proportion to the verification buffer size.

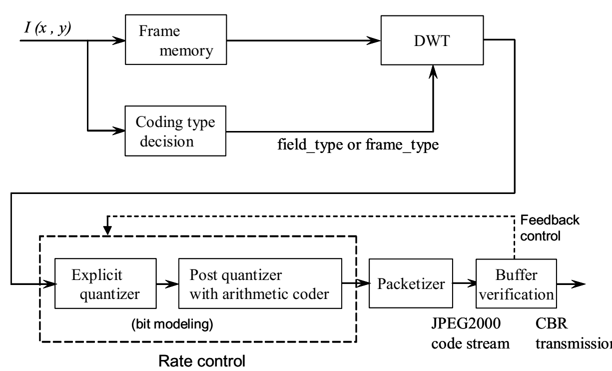


Figure 2.1 Block diagram of assumed encoder.

3. NEW CODING TECHNOLOGIES

3.1. OPTIMAL RATE CONTROL

As a rate control scheme constituting JPEG2000 encoder, an R-D (Rate-Distortion) optimized post quantization is generally adopted. By applying the scheme, MSE of a coded picture is certainly kept at minimum value. However, the scheme allows the number of truncated coding passes to become different among adjacent code blocks in a sub-band, whereby truncating one coding pass results in doubling the quantization step size for the corresponding code block. The inconsistency in the spatial distribution of quantization error might cause unexpected subjective picture quality degradation, while this problem is more apparent for motion pictures rather than still pictures. From this standpoint, we investigated a modified rate control scheme under following assumptions.

- Our rate control is conducted in principle along a hybrid process of an explicit quantizer and a post quantizer as shown in Figure 2.1. When encoding a picture, the target number of coded bits is assigned first then the rate control in order to satisfy the target is conducted using a series of quantization process.
- In an explicit quantizer, Q-table is defined by the matrix *delta* as shown in equation (3.1). In the equation, *dqt* indicates the default Q-table referred to as an initial value of *delta*, and *cq* is a weighting factor commonly multiplied for every component of *dqt* which is updated picture by picture from the coding result.

$$delta[band] = dqt[band] \times cq \tag{3.1}$$

- In a post quantizer, the number of truncated coding passes is kept identical over all code blocks constituting a picture in order to keep a uniformly distributed quantization error over code blocks.
- The fluctuation in the number of coded bits for each picture is eliminated in order to minimize the buffer size.

The detailed rate control procedure is described in the following.

(1) Decision of target number of coded bits

When starting the encoding process for a current picture, the decision of the target number of coded bits for the picture is required. In our proposed scheme, *R(t)*, the number of bits available for the t th picture is calculated with the equation (3.2) which compensates for the difference between the actual value and the target value with respect to the verification buffer occupancy.

$$\begin{cases} R(t) = \{v_0 + (L_w + t)\times B - (v_T + \sum_{i=0}^{t-1} S(i))\} / L_w \\ B = bit\_rate / picture\_rate \end{cases} \tag{3.2}$$

L_w: Window size of rate control

v_0: Buffer occupancy when the first picture is extracted from the verification buffer.

v_T: Target value for verification buffer occupancy.

$S(i)$: Actual number of encoded bits for i th picture.

(2) Post quantization

In the post quantization process, *np*, the number of truncated coding passes is adjusted to be identical among every code block in a picture. The value *np* which can minimize $|r(t, np) - R(t)|$ is selected as an optimal value, where $r(t, np)$ indicates the total number of coded bits under *np* coding passes which are truncated.

(3) Updating Q-table

After encoding a current picture, the weighting factor *cq* is updated for the next picture by equation (3.3) where f indicates a non-linear function derived to approximate the relationship between x and y. The actual samples of (x, y) are obtained by equation (3.4) from the coding result of a current picture. In case $r(t,0)$ becomes lower than $R(t)$, the number of coded bits must be insufficient even when all coding passes are accepted. To overcome this problem, an updating algorithm is designed as the number of coded bits under np=3 to be kept close to $R(t)$.

$$\begin{cases} cq(t+1) = cq(t) \times \frac{x_T}{2} \\ f(x_T) = R(t+1) \end{cases} \tag{3.3}$$

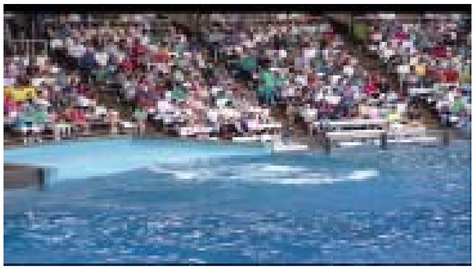
$$\begin{cases} y = f(x) \\ x = 2^{\frac{np}{3}} \text{ , } y = r(t, np) \end{cases} \tag{3.4}$$

The proposed rate control technology is applicable even for JPEG2000 coding with a tile structure. Under a tile structure, an encoded image area is divided into several small tiles while the JPEG2000 coding process is applied for each tile independently. Yet the assignment of target number of bits and the determination of the Q-table can be conducted at a picture level before starting the tile level coding process. The picture level rate control is identical with the approach introduced above. The obtained Q-table is commonly applied for every tile in a picture to eliminate these tile artifacts. At the beginning of tile level coding, the target number of bits is decided for each tile then a post quantization is independently applied for each tile. Therefore, $r_j(t, 3)$ which corresponds to the contribution of j th tile to $r(t, 3)$ is additionally defined for tile structures. For j th tile, the target number of bits $R_j(t)$ can be determined from the value of $R(t)$ as shown in equation (3.5) where *tile_cnt* indicates the number of tiles in a picture. The similar post quantization approach described for no tiling can be applied for each tile respectively.

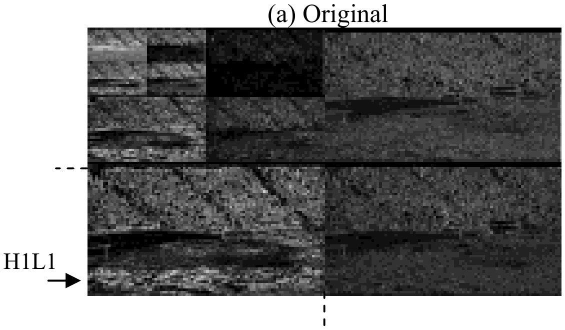
$$\begin{cases} R_j(t) = R(t) \times \frac{r_j(t-1, 3)}{r(t-1, 3)} \\ r(t, 3) = \sum_{j=0}^{tile\_cnt-1} r_j(t, 3) \end{cases} \tag{3.5}$$

3.2. SCENE ADAPTIVE CODING TYPE SELECTION

In a JPEG2000 encoder for interlaced TV signals, a frame and field type can be selected as the JPEG2000 coding type. The selection of the coding type depends on the coding characteristics of the input sequence. In general, frame type coding is statically assigned because of its high coding efficiency. Yet frame type coding handles two fields whose presentation time differs by one field time in an interlaced frame. Frame type coding also generates horizontal comb-tooth regions along the object contours in sequences with high speed motion objects or camera operations [2]. The comb-tooth region generates a corresponding signal in the H1L1 sub-band as shown in Figure 3.1, and the total coding efficiency deteriorates because of additional bit allocations for signals needed to suppress the subjective quality distortion around the moving object contours.



(a) Original



(b) Sub-band decomposition

Figure 3.1 Example of sub-band decomposition of interlaced HDTV signals ("Whale show ": 1080i test sequence)

Thus we developed and proposed an optimum scheme for coding type selection of each interlaced frame. In this study, each interlaced frame is first fed into a coding type decision unit to assign an optimal coding type to the JPEG2000 coding unit then the corresponding frame is encoded under the assigned coding type as shown in Figure 2.1. In the proposed scheme, a coding type decision is conducted along equations (3.6) - (3.9), where $I(x, y)$ indicates the input luminance value of the pixel (x, y) ($0 \leq x \leq hsize$-1, $0 \leq y \leq vsize$-1) in a current interlaced frame. In principle, the proposed scheme estimates the correlation in the vertical direction for each candidate with low computational cost, and a weighting factor $a(R)$ (> 1.0) which is a constant value dependent on the coding bit-rate R is introduced into equation (3.8) to account for differences in the amount of coded side information between two candidates.

In the practical implementation of a JPEG2000 encoder, frequency weighting control among sub-bands is applied to achieve a subjective picture quality improvement. In frame type coding, the comb-tooth region is concentrated in the H1L1 sub-band whose priority decreases as a result of the general visual weighting control. Coding type selection should therefore depend on the visual weighting factor assigned for the H1L1 sub-band. In our proposed scheme, the coefficient $a(R)$ is replaced by a function $a(R, w_v)$ in equation (3.8) only when a visual weighting control is active. The function $a(R, w_v)$ calculates the weighting factor by equation (3.9) where w_v indicates the visual weighting factor of the H1L1 sub-band. In the equation, $b(R)$ indicates a constant value which depends on the coding bit-rate.

$$c_I = \sum_{y=0}^{vsize-3} \sum_{x=0}^{hsize-1} |(I(x,y) - I(x,y+2)| \quad (3.6)$$

$$c_P = \sum_{y=0}^{vsize-3} \sum_{x=0}^{hsize-1} |(I(x,y) - I(x,y+1)| \quad (3.7)$$

i) $c_P < a(R) \times c_I$ (3.8)

frame_type is selected

ii) otherwise

field_type is selected

$$a(R, w_v) = b(R) \times \log_2 w_v + a(R) \quad (3.9)$$

4. EXPERIMENTAL RESULTS

To evaluate the coding picture obtained with the proposed HDTV coding scheme, coding experiments were performed under the encoding conditions shown in Table 4.1. In these experiments, proposed technologies were implemented on the Motion JPEG2000 Verification Model [3].

Table 4.1 Coding experiment conditions.

Input sequence	1920x1080 / 59.94Hz (interlace) 4:2:2 digital component (8bit)
Coding bit-rate	50Mbps (0.8bit/pixel)
DWT filter	(9,7) irreversible
Sub-band level	3 levels
Code block size	64 x 64
Frequency weighting	None
Tile size	No tiling

In order to evaluate the stability of the proposed rate control, verification buffer occupancy value is observed as shown in Figure 4.1 and Figure 4.2. The vertical axis indicates the percentage of occupancy value normalized by the averaged number of encoded bits for a picture. In this experiment, field type coding was assumed as a coding type, and the performance under the tile structure whose size is 1920 x 135 was also evaluated as a result labeled with "*tile_cnt*=4" in those figures. From the results, the fluctuation in verification buffer occupancy is less than ±8%. With regards to the required buffer size, only a 10% increase seems sufficient compared to the minimum cases where the number of encoded bits can be kept identical for every picture. Therefore, the proposed scheme was confirmed to be very effective in eliminating codec delay. From further experiments, it was also confirmed that the proposed scheme did not cause any degradation in PSNR compared to the minimum case described above whose results were obtained by manually setting the proper Q-table under which the number of consumed bits is kept identical for every picture.

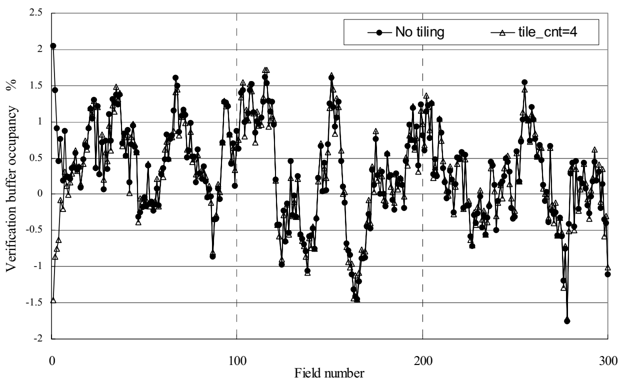


Figure 4.1 Verification buffer occupancy result obtained by proposed scheme (European market).

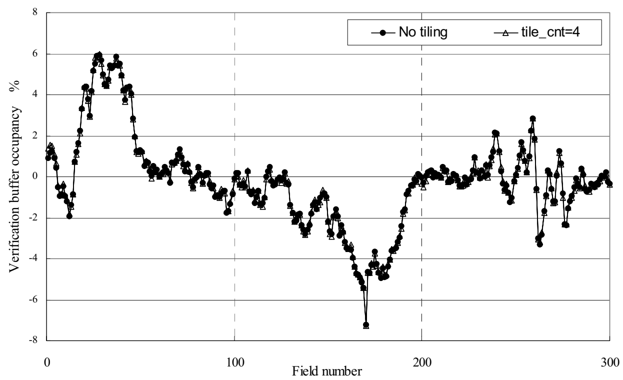


Figure 4.2 Verification buffer occupancy result obtained by proposed scheme (Whale show).

The PSNR results obtained when applying the adaptive coding type selection under the proposed rate control are shown in Figure 4.3 and Figure 4.4. In order to evaluate the coding gain compared to the conventional scheme, the results obtained under the static coding type assignment are also shown for frame type and field type in those figures respectively. These figures include two results respectively. One is the result obtained under the typical frequency weighting control specified in the JPEG2000 IS annex J [4]. The other is the result obtained without frequency weighting.

We first confirmed that our key technology caused no picture quality degradation. We also found that picture quality was improved in almost every frame constituting a sequence as a result of selecting a more efficient coding type from the two candidates. The gain in PSNR achieved compared to static assignment of the frame type corresponds to the improvement in input signals with the comb-tooth regions which had caused problems in conventional schemes, while the maximum gain was extended from 0.5dB to 1.0dB depending on the sequence.

4. CONCLUSIONS

In this paper, we described advanced key technologies that can be incorporated into highly compressed HDTV coding

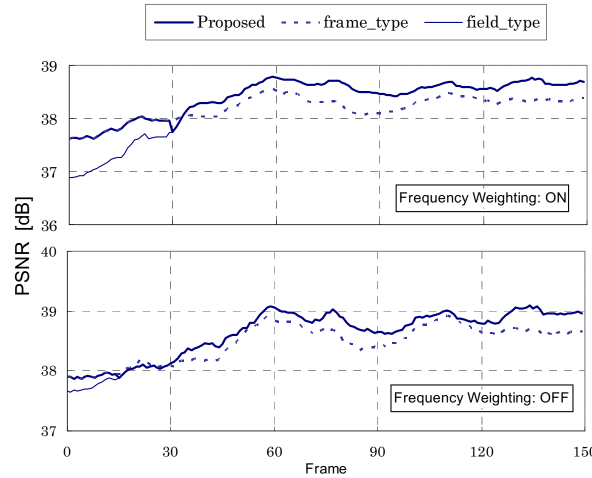


Figure 4.3 PSNR result obtained by proposed scheme (European Market)

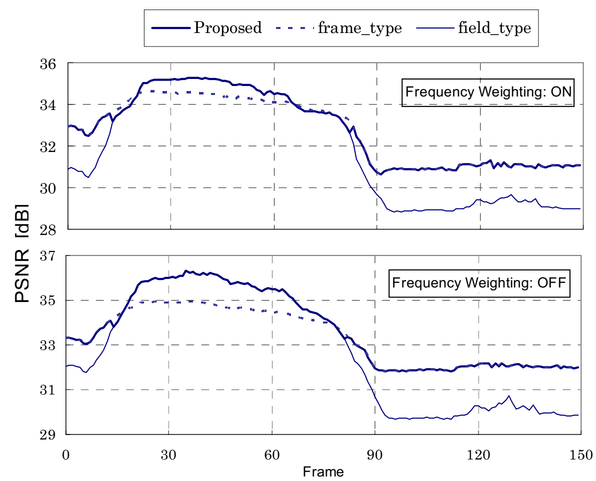


Figure 4.4 PSNR result obtained by proposed scheme (Whale show)

based on Motion JPEG2000. The results from coding experiments confirmed that additional coding gains could be achieved by applying our proposed technologies for JPEG2000 base HDTV coding. We also showed that our HDTV coding scheme was extremely effective in overcoming the inherent problems that accompany handling interlaced TV signals by conventional schemes.

5. REFERENCES

[1] ISO/IEC JTC1/SC29/WG1 N2250, "Motion JPEG2000 Final Draft International Standard 1.0", Sept. 2001.
[2] T. Kuge, "Wavelet Picture Coding and Its Several Problems of the Application to the Interlace HDTV and the Ultra-high Definition Images", ICIP2002, Vol. III, pp.217-220, Sept. 2002.
[3] ISO/IEC JTC1/SC29 WG1 N2144, "Motion JPEG2000 Verification Model ver.4.1", June 2001.
[4] ISO/IEC JTC1/SC29/WG1 N1890, "JPEG2000 part I Final Draft International Standard", Sept. 2000.