RATE-DISTORTION OPTIMIZED IMAGE CODING ALLOWING LOSSLESS CONVERSION TO JPEG COMPLIANT BITSTREAMS

Ichiro Matsuda, Shunya Hashimoto, Hiroyuki Koike, Hisashi Aomori and Susumu Itoh

Department of Electrical Engineering, Faculty of Science and Technology,
Tokyo University of Science

2641 Yamazaki, Noda-shi, Chiba 278-8510, JAPAN
phone: +81 4 7124 1501 (ext.3740), fax: +81 4 7124 9367
email: matsuda@ee.noda.tus.ac.jp

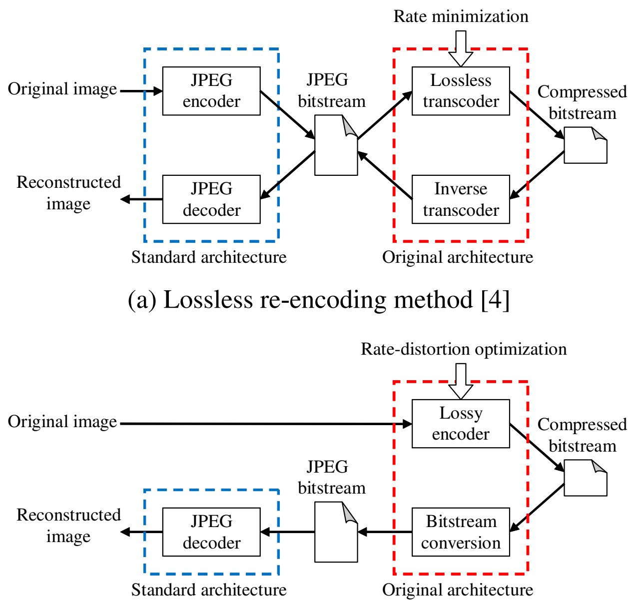
ABSTRACT

This paper proposes an efficient lossy image coding scheme which has a sort of compatibility with the JPEG standard. In this scheme, our encoder directly compresses an image into a specific bitstream. On the other hand, the reconstructed image is given by the standard JPEG decoder with a help of lossless bitstream conversion from the specific bitstream to the JPEG compliant one. This asymmetric structure of encoding and decoding procedures enables us to utilize modern coding techniques such as intra prediction, rate-distortion optimization and arithmetic coding in encoding process, while allowing almost all image manipulation software to directly import image contents without any loss of quality. In other words, the proposed scheme utilizes the JPEG compliant bitstream as an intermediate format for the decoding process. Simulation results indicate that the rate-distortion performance of the proposed scheme is better than those of the current state-of-the-art lossy coding schemes.

1. INTRODUCTION

The JPEG baseline system [1] has been the most popular international standard for still image coding over the past two decades and almost all image applications support the JPEG standard for importing and exporting image contents. Though more efficient and functional coding schemes such as JPEG 2000 [2] and JPEG XR [3] have been standardized, they are suffering from lack of the applications which support their image format. This is one reason why the old JPEG standard is even now widely used as a common image format and the succeeding standards have not yet been popular.

From this point of view, we previously proposed a lossless re-encoding method which compresses JPEG images without any loss of quality [4]. As illustrated in Figure 1 (a), the method can be seen as a kind of transcoding scheme where conversion between the JPEG bitstream and the compressed one is reversible. Lossless compression of the JPEG image is realized by extracting quantized discrete cosine transform (DCT) coefficients from the given bitstream and re-encoding them without additional quantization. In this framework, we can utilize modern coding techniques for redundancy removal and efficient entropy coding to reduce the amount of the output bitstreams. Practically, our method employed the techniques of H.264-like intra prediction and arithmetic coding, and as a result, archived 18-28 % bitrate savings against the original JPEG bitstreams [4]. This type of lossless re-encoding method is useful for the images whose original information captured by image sensors was



(a) Lossless re-encoding method [4]

(b) Proposed scheme

Figure 1: Structures of encoding and decoding procedures.

lost when they were encoded by the lossy algorithm of the JPEG standard. However, in the situation where the original information (i.e. original image) is available, the lossless re-encoding method generally attains poor rate-distortion performance compared with the state-of-the-art lossy coding schemes due to constraints of coding syntax determined by the JPEG encoder.

To cope with this problem, in this paper, we propose an efficient lossy coding scheme which provides compatibility with the JPEG standard in a certain sense. The scheme directly encodes the original image into a specific bitstream like the conventional lossy coding schemes. However, decoding process is carried out by the standard JPEG decoder with a help of lossless bitstream conversion as show in Figure 1 (b). In this case, a distortion between the original and reconstructed images is controllable in our encoder. Consequently, not only quantization step sizes but also several coding parameters, such as intra prediction modes and probability models used for the arithmetic coding, can be optimized in a rate-distortion sense. This type of coding scheme could be a practical solution for satisfying both requirements of high coding performance and portability of the compressed image contents.

2. BLOCK-ADAPTIVE INTRA PREDICTION

The JPEG baseline system adopts a DCT-based coding algorithm where image signals are transformed, quantized and entropy coded on a block-by-block basis. In this algorithm, inter-block correlations of the image signals are not taken into account with the exception of a DC component of the DCT coefficients. To overcome this weakness, the proposed scheme employs a block-adaptive intra prediction technique which was developed for the H.264/AVC video coding standard [5]. Since the size of DCT used in the JPEG standard is fixed to 8×8 pels, the procedure of '8×8 intra prediction' specified in High profile of the H.264/AVC standard is performed to obtain a predicted image $\hat{x}(i, j)$ $(i, j = 0, 1, \ldots, 7)$ in each block. Figure 2 illustrates positions of pels used for the '8×8 intra prediction' as shaded boxes.

It should be noted, in our scheme, that the reconstructed image $x'(i, j)$ needs to have consistency with the JPEG standard. In other words, DCT coefficients correspond to the $x'(i, j)$ must be representable by the inverse quantizer used in the JPEG standard. To meet this requirement, a quantization process is carried out before the prediction and the prediction residuals are calculated in DCT domain:

$$r(i, j) = Q\left[y(i, j), \Delta q(i, j)\right] - \hat{y}(i, j)/\Delta q(i, j), \qquad (1)$$

where $Q[\,\cdot\,, \Delta q]$ means a JPEG compliant quantizer with a step size of Δq, $y(i, j)$ and $\hat{y}(i, j)$ are DCT coefficients of the original and predicted images, respectively. A matrix $\Delta q(i, j)$ containing the quantizer step sizes is called the quantization table and its elements are changeable for each color component in the JPEG standard. In general, the prediction residual $r(i, j)$ becomes a real number. However, it can be encoded losslessly without additional quantization because output of the quantizer $Q[\,\cdot\,, \Delta q]$ in Eq.(1) is an integer. The best prediction mode $d \in \{1, 2, \ldots, 9\}$ from among eight directional prediction modes and a DC prediction mode specified in the '8×8 intra prediction' [6] is chosen for each block so that a coding rate of the prediction residuals can be a minimum. The detailed optimization procedure will be described later.

3. CODING OF PREDICTION RESIDUALS

It is known that a probability distribution function (PDF) of two-dimensional DCT coefficients obtained from natural images can be modeled by the following generalized Gaussian function [7]:

$$f(y \mid \mu, \sigma, c) = \frac{c \cdot \eta(\sigma, c)}{2\Gamma(1/c)} \cdot \exp\left\{-\eta(\sigma, c) \cdot \left|y - \mu\right|^{c}\right\}, \qquad (2)$$

$$\eta(\sigma, c) = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/c)}{\Gamma(1/c)}}, \qquad (3)$$

where $\Gamma(\cdot)$ is the gamma function, μ and σ are the mean value and the standard deviation of the DCT coefficients, respectively. In addition, c represents a shape parameter which controls sharpness of the function. In this paper, it is assumed that the above PDF model is also applicable to the prediction residuals calculated by Eq.(1). In general, however, the standard deviation σ and the shape parameter c depend not only on frequency indices (i, j) but also on local activity of the image signals. In order to capture such a statistical property, we classify all the blocks into M classes and prepare two kinds of look-up tables $\sigma_m(i, j)$ and $c_m(i, j)$, which assign appropriate values of the standard deviation and the shape parameter to the respective frequencies, for each class $(m = 1, 2, \ldots, M)$. On the other hand, the mean value of the prediction residual tends to be zero as a result of the block-adaptive intra prediction. According to these considerations, a conditional probability of occurrence of the quantization output $q = Q[y(i, j), \Delta q] \in \mathbb{Z}$, when the current block belongs to the m-th class and a DCT coefficient $\hat{y}(i, j)$ of the predicted image $\hat{x}(i, j)$ is given, is estimated by:

$$\Pr_{i,j}(q \mid m, \hat{y}(i, j)) = \int_{q-0.5}^{q+0.5} f(y \mid \mu, \sigma, c)\, \mathrm{d}y, \qquad (4)$$

$$\mu = \hat{y}(i, j)/\Delta q(i, j), \qquad (5)$$

$$\sigma = \sigma_m(i, j), \qquad (6)$$

$$c = c_m(i, j). \qquad (7)$$

These equations indicate that the probability is calculated by shifting a center of the generalized Gaussian function to the predicted position and integrating it between quantization thresholds with a uniform interval as shown in Figure 3. The probabilities calculated for all possible quantization levels in this way are used for entropy coding of the actual prediction residual $r(i, j)$. In our implementation, the rangecoder [8] which is known as a fast multisymbol arithmetic coder is employed for the entropy coding. Figure 4 illustrates the overall encoding procedures in the proposed scheme.

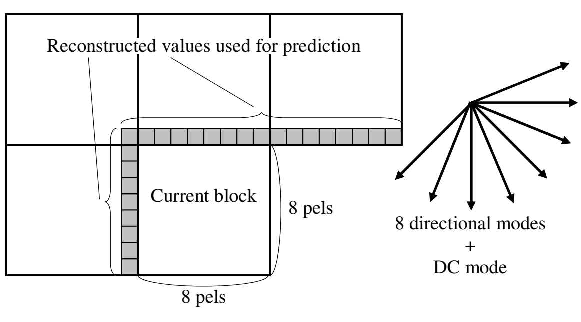


Figure 2: Block-adaptive intra prediction.

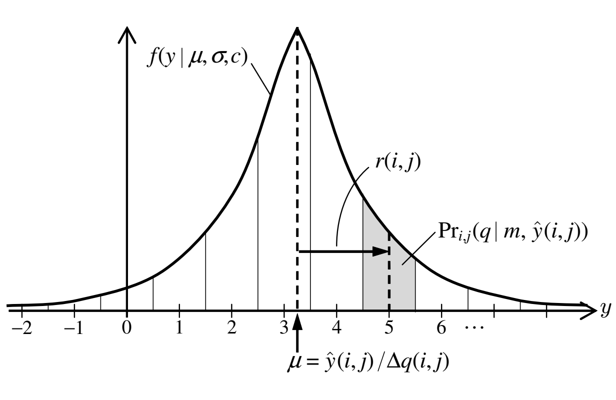


Figure 3: Occurrence probability of a quantized DCT coefficient ($q = 5$).

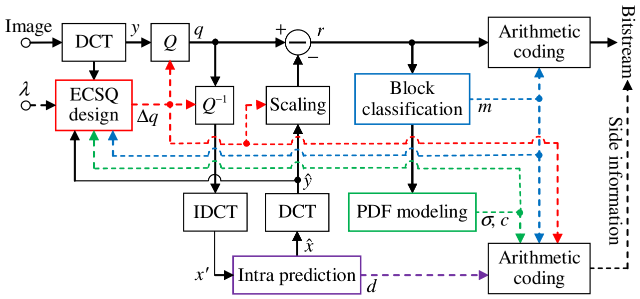


Figure 4: Block diagram of the proposed encoder.

4. RATE-DISTORTION OPTIMIZATION OF CODING PARAMETERS

Table 1 lists coding parameters which are encoded as side information for the proposed scheme. In the case of the lossless re-encoding method [4], the quantization table $\Delta q(i, j)$ was determined when the original image was encoded by the JPEG standard encoder and fixed throughout the re-encoding process. Therefore, a distortion between the original and reconstructed images was unchangeable and the optimization of coding parameters was considered as a rate minimization problem. In contrast to that case, the proposed scheme has a freedom to control the distortion by changing the quantization table and/or a quantization method. This allows us to take advantages of a more flexible coding strategy based on rate-distortion optimization. The optimization procedure of the proposed scheme is divided into two stages: the first stage is dedicated to the optimal quantizer design while the second one optimizes the remaining parameters. Both stages described below are repeated alternately until all of the parameters converge.

4.1 Optimization of quantization step sizes with ECSQ

In this stage, we want to minimize the following Lagrangian cost function by tuning the quantization step sizes:

$$J(\lambda) = D + \lambda \cdot R, \tag{8}$$

where D is a distortion defined as the mean square error (MSE) between the original and reconstructed images and R represents the coding rate required for the reconstructed image. In the proposed scheme, a positive constant λ called the Lagrangian multiplier is used to control rate-distortion trade off. As a matter of fact, this kind of rate-distortion optimization problem was already studied in the context of JPEG-compatible encoding [9, 10]. Since the JPEG standard uses the run-length encoding method for the quantized DCT coefficients, a rather complicated technique such as the dynamic programming method is needed to solve the problem [10]. On the other hand, the proposed scheme encodes the DCT coefficients one-by-one using the arithmetic coding. Therefore, the distortion $D_{i,j}(\Delta q)$ and the coding rate $R_{i,j}(\Delta q)$ with respect to the DCT coefficients quantized with a step size Δq can be independently measured for each frequency:

$$D_{i,j}(\Delta q) = \sum_{\text{for all blocks}} \left\{ y(i,j) - \Delta q \cdot Q[y(i,j), \Delta q] \right\}^2, \tag{9}$$

$$R_{i,j}(\Delta q) = -\sum_{\text{for all blocks}} \log_2 \Pr_{i,j}\Big(Q[y(i,j), \Delta q] \,\Big|\, m, \hat{y}(i,j) \Big). \tag{10}$$

This means that an individual quantization step size can be determined by solving the following sub-optimization problem:

$$\Delta q(i,j) = \mathop{\mathrm{argmin}}_{\Delta q \in \{1,2,\ldots,255\}} \Big\{ D_{i,j}(\Delta q) + \lambda \cdot R_{i,j}(\Delta q) \Big\}. \tag{11}$$

Strictly speaking, a predicted value $\hat{y}(i, j)$ in Eq.(10) depends on quantization step sizes of other frequencies because it is obtained as a result of the block-adaptive intra prediction and DCT. Therefore, we start the optimization procedure with initial values of $\Delta q(i, j) = \lfloor 2\sqrt{\lambda/0.85} + 0.5 \rfloor$ $(i, j = 0, 1, \ldots, 7)$, and then gradually refine them in each iteration.

Furthermore, we can also change the quantization method, or quantization thresholds, as far as the representation levels of the quantizer conform to the JPEG standard. Thereupon, we introduce a technique of the entropy-constrained scalar quantizer (ECSQ) [11] into the optimization process. The ECSQ used in the proposed scheme is defined by:

$$Q[y, \Delta q] = \mathop{\mathrm{argmin}}_{q \in \mathbb{Z}} \Big\{ (y - \Delta q \cdot q)^2 - \lambda \cdot \log_2 \Pr_{i,j}(q \mid m, \hat{y}(i,j)) \Big\}. \tag{12}$$

4.2 Optimization of other parameters

Once the quantization step sizes of the ECSQ are determined, the distortion of the reconstructed image is fixed and the first term of the right hand side of Eq.(8) can be ignored. Therefore, optimization of the PDF models ($\sigma_m(i, j)$ and $c_m(i, j)$) for each class, as well as selection of the prediction modes (d) and the class labels (m) for each block, is conducted so that the overall coding rate can be a minimum. Moreover, the number of classes (M), which is initially set to $M = 64$, is gradually decreased within this optimization stage to finally obtain its appropriate setting [4]. In this paper, the parameters $\sigma_m(i, j)$ and $c_m(i, j)$ are chosen from the following given sets of values:

$$\sigma_m(i,j) \in \{0.1 \cdot (5/4)^n \mid n = 0, 1, \ldots, 31\}, \tag{13}$$

$$c_m(i,j) \in \{0.5, 1.0, 1.5, 2.0\}. \tag{14}$$

Table 1: Parameters required for the proposed scheme.

Description	Number of parameters	Allowed values
Quantization table $\Delta q(i, j)$	$8 \times 8 \times N_q$	$1, 2, \ldots, 255$
Standard deviation $\sigma_m(i, j)$	$8 \times 8 \times M$	See Eq.(13)
Shape parameter $c_m(i, j)$	$8 \times 8 \times M$	See Eq.(14)
Prediction mode d	N_b	$1, 2, \ldots, 9$
Classification label m	N_b	$1, 2, \ldots, M$

N_q: The number of quantization tables ($N_q = 1$ for monochrome images)
N_b: The number of 8×8 blocks

5. EXPERIMENTAL RESULTS

In order to evaluate coding performance of the proposed scheme, computer simulations are conducted using several

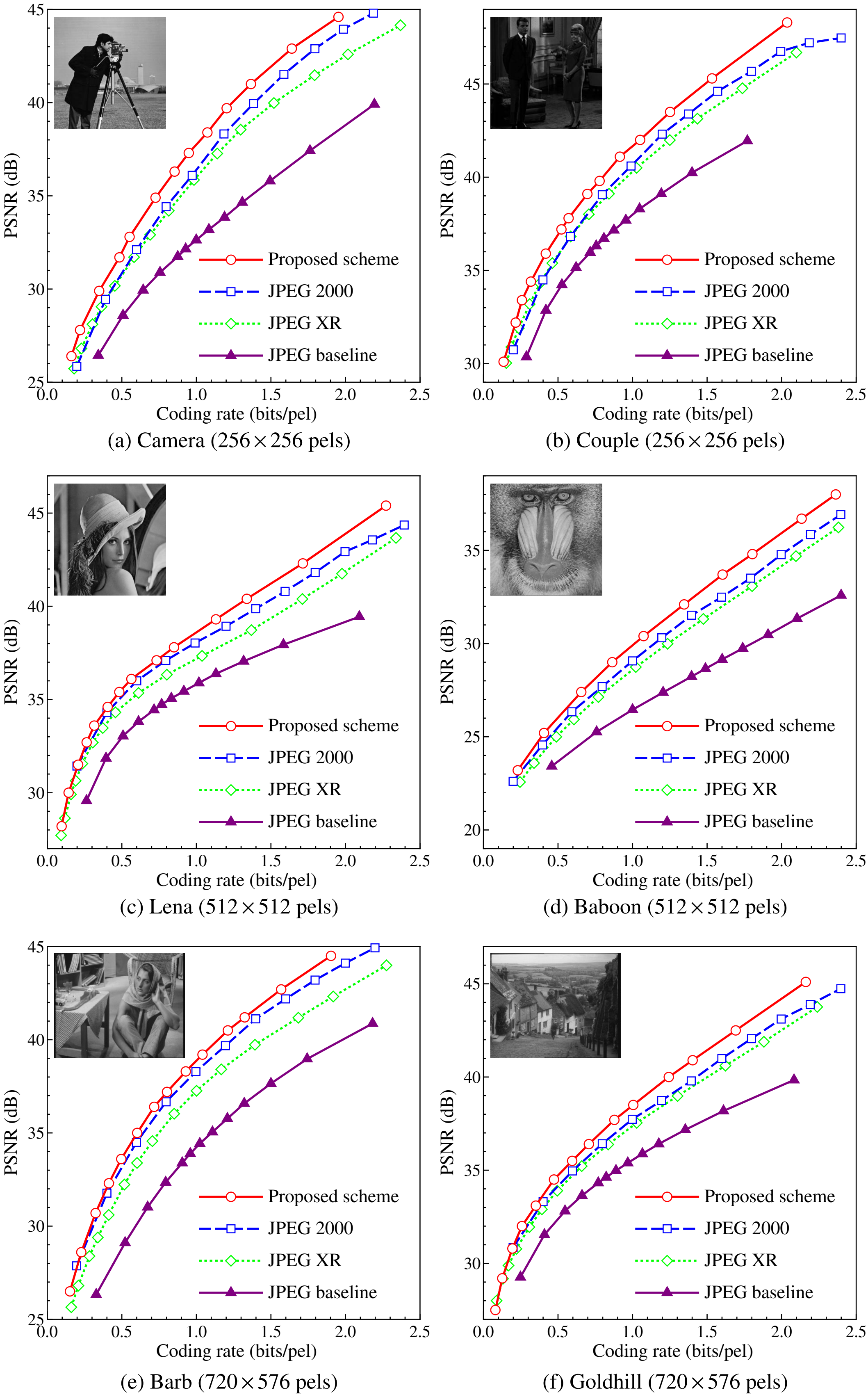


Figure 5: Coding performance.

14	12	14	14	15	15	15	14
14	13	15	15	15	14	16	15
14	14	14	15	14	15	15	17
14	14	14	14	15	16	17	16
14	15	15	14	16	14	14	14
14	15	15	13	15	14	14	14
16	16	15	15	17	15	15	15
16	17	16	16	16	16	16	18

(a) Proposed scheme
0.95 bits/pel, 37.3 dB

5	3	3	5	7	12	15	18
4	4	4	6	8	17	18	17
4	4	5	7	12	17	21	17
4	5	7	9	15	26	24	19
5	7	11	17	20	33	31	23
7	11	17	19	24	31	34	28
15	19	23	26	31	36	36	30
22	28	29	29	34	30	31	30

(b) JPEG baseline
1.76 bits/pel, 37.4 dB

Figure 6: Examples of the quantization tables (Camera).

monochrome images. Figure 5 plots rate-PSNR curves of the proposed scheme together with those of the standard coding schemes: JPEG 2000 (JasPer version 1.900.1 [12], 9/7 wavelet), JPEG XR (Reference software version 1.8 [13]) and JPEG baseline (IJG library version 6b [14]). Coding rates of the proposed scheme include all of the necessary side information listed in Table 1. It is demonstrated that the proposed scheme achieves better coding performance than the current standard schemes over a wide range of coding rates. The PSNR gain is up to 1.2 dB compared with the JPEG 2000 standard which is considered as the most efficient coding standard for still images. Figure 6 compares the quantization tables used in the proposed scheme and the JPEG baseline. In general, the JPEG baseline encoder uses a scaled version of the default quantization table shown in the Annex of the JPEG standard [1]. This default quantization table was determined based on the psychovisual test, and therefore not intended to give better rate-distortion performance. On the other hand, the quantization table used in the proposed scheme has relatively flat values. It is consistent with the classical optimum bit allocation theory which states that variances of quantization errors should be equalized for all frequencies [15]. Small variation in the quantization table seem to result from different properties of the PDFs at the respective frequencies. If we want to improve subjective quality of the reconstructed images, it is possible to use other distortion metrics based on human visual perception. For this purpose, some sort of the WMSE metric calculated by weighting quantization errors of DCT coefficients [16] is suitable since the distortion is always measured in DCT domain in the proposed scheme.

6. CONCLUSIONS

We have proposed a new type of lossy image coding scheme which has an asymmetric structure in encoding and decoding processes. In the scheme, the encoding process is carried out by our single encoder while the decoding process utilizes the standard JPEG decoder together with the lossless bitstream converter from the encoded bitstream to JPEG compliant one. Such a coding structure is beneficial to most of the existing image manipulation software since they can import image contents without any quality loss using their own built-in JPEG decoder. Simulation results indicate that, although the proposed scheme has a sort of compatibility with the old JPEG standard, it attains better coding performance than the newer standards such as JPEG 2000.

REFERENCES

[1] ITU-T Rec. T.81 | ISO/IEC 10918-1, "Information Technology — Digital Compression and Coding of Continuous-Tone Still Images: Requirements and Guidelines," 1994.

[2] ITU-T Rec. T.800 | ISO/IEC 15444-1, "Information Technology — JPEG 2000 Image Coding System — Part 1: Core Coding System," 2001.

[3] ITU-T Rec. T.832 | ISO/IEC 29199-2, "Information Technology — JPEG XR Image Coding System Part 2: Image Coding Specification," 2009.

[4] I. Matsuda, Y. Nomoto, K. Wakabayashi and S. Itoh, "Lossless Re-encoding of JPEG Images Using Block-Adaptive Intra Prediction," *Proc. of 16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, L3-6, Aug. 25–29, 2008.

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

[6] D. Marpe, T. Wiegand and S. Gordon, "H.264/MPEG4-AVC Fidelity Range Extensions: Tools, Profiles, Performance, and Application Areas," *Proc. of 2005 IEEE International Conference on Image Processing (ICIP 2005)*, Genova, Italy, vol.1, pp.593–596, Sep. 11-14, 2005.

[7] F. Müller, "Distribution Shape of Two-Dimensional DCT Coefficients of Natural Images," *Electronics Letters*, vol.29, no.22, pp.1935–1936, Oct. 1993.

[8] M. Schindler, "A Fast Renormalization for Arithmetic Coding," *Proc. of Data Compression Conference (DCC '98)*, Snowbird, Utah, p.572, Mar. 30–Apr. 1, 1998.

[9] S. W. Wu and A. Gersho, "Rate-Constrained Picture-Adaptive Quantization for JPEG Baseline Coders," *Proc. of 1993 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-93)*, Minneapolis, Minnesota, vol.5, pp.389–392, Apr. 27-30, 1993.

[10] M. Crouse and K. Ramchandran, "Joint Thresholding and Quantizer Selection for Transform Image Coding: Entropy-Constrained Analysis and Applications to Baseline JPEG," *IEEE Trans. on Image Processing*, vol.6, no.2, pp.285–297, Feb. 1997.

[11] G. J. Sullivan, "Efficient Scalar Quantization of Exponential and Laplacian Random Variables," *IEEE Trans. on Information Theory*, vol.42, no.5, pp.1365–1374, Sep. 1996.

[12] http://www.ece.uvic.ca/~mdadams/jasper/

[13] ITU-T Rec. T.835 | ISO/IEC 29199-5, "Information Technology — JPEG XR Image Coding System Part 5: Reference Software," 2010.

[14] http://www.ijg.org/

[15] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Englewood Cliffs, New Jersey 07632: Prentice-Hall, 1984.

[16] B. Macq, "Weighted Optimum Bit Allocations to Orthogonal Transforms for Picture Coding," *IEEE J. on Selected Areas in Commun.*, vol.10, no.5, pp.875–883, June 1992.