COEFFICIENT-WISE INTRA PREDICTION FOR DCT-BASED IMAGE CODING

Ichiro Matsuda, Yusuke Kameda and Susumu Itoh

Department of Electrical Engineering, Faculty of Science and Technology,
Tokyo University of Science
2641 Yamazaki, Noda-shi, Chiba 278-8510, JAPAN
email: matsuda@ee.noda.tus.ac.jp

ABSTRACT

This paper proposes an adaptive intra prediction method for DCT-based image coding. In this method, predicted values in each block are generated in spatial domain like the conventional intra prediction methods. On the other hand, prediction residuals to be encoded are separately calculated in DCT domain, i.e. differences between the original and predicted values are calculated after performing DCT. Such a prediction framework allows us to change the coding process from block-wise order to coefficient-wise one. When the coefficient-wise order is adopted, a block to be predicted is almost always surrounded by partially reconstructed image signals, and therefore, efficient interpolative prediction can be performed. Simulation results indicate that the proposed method is beneficial for removing inter-block correlations of high-frequency components.

***Index Terms*—** Image coding, DCT, intra prediction, progressive JPEG, linear interpolation

1. INTRODUCTION

Discrete Cosine Transform (DCT) has been widely used as the main component of image and video coding systems. In such systems, two dimensional DCT is usually performed on a block-by-block basis to decorrelate image signals within the respective blocks. However, DCT itself has no effect on inter-block correlations which may exist among adjacent blocks. In order to compensate this limitation, traditional coding schemes such as JPEG [1] employ a simple predictive coding technique called Differential Pulse-Code Modulation (DPCM) only for a DC component of DCT coefficients. In H.263+ [2] and MPEG-4 [3], the DPCM is also applied to a few AC components in an adaptive way. Since these techniques are directly performed in DCT domain, it is generally difficult to predict directional textures which may be decomposed into multiple DCT coefficients. To cope with this problem, H.264/AVC [4] adopts a so-called intra prediction method which is carried out in spatial domain, i.e. prior to performing DCT. This kind of prediction provides a high degree of freedom in exploiting the inter-block correlations: up to nine intra prediction modes most of which are designed for directional textures are adaptively selected block-by-block in the H.264/AVC. After the success of the H.264/AVC, an advanced intra prediction method called angular prediction which supports a total of 33 prediction directions is developed and adopted for new H.265/HEVC video coding standard [5]. In both of the intra prediction methods, coding processes including prediction, calculation of prediction residuals, transform, quantization and entropy coding are basically performed in block-wise order. Consequently, reference pels used for the prediction are always located in fixed causal blocks: they are at upper and left sides of the current block if the coding processes are done in raster-scan order. As a result, the conventional intra prediction methods are considered as extrapolation of already encoded image signals in a spatial direction.

In the case of inter-frame video coding, it is known that coding efficiency can be improved by changing coding order of frames. This is because we can make use of the efficient bi-directional prediction, i.e. interpolative prediction in a temporal direction [6]. Motivated by this fact, a spatial version of the bi-directional prediction method was proposed for intra-frame video coding [7]. In this method, two kinds of coding order were adaptively switched within a 2×2 array of blocks and the bi-directional prediction was realized by weighting two predicted values extrapolated from different directions. In [8], we enhanced the method by increasing variation of prediction modes as well as that of possible block coding orders. As a result, it was found that though the interpolative prediction in a spatial direction is evidently beneficial, there are several drawbacks in changing coding order of blocks. First, it increases the computational complexity for testing all of the possible coding orders at the encoder side. Second, it requires extra side information to indicate the resulting coding order. Furthermore, in exchange of taking advantage of the interpolative prediction, some blocks should accept poor performance of the extrapolative prediction using reference pels located apart from the current block. According to our research, no more than 25 % of the pels can benefit form the interpolative prediction even by adapting the coding order [8].

In this paper, we modify the procedure of intra prediction to calculate prediction residuals in DCT domain. Such a modification allows us to change the coding process from block-wise order to coefficient-wise one. When the coefficient-wise order is adopted, the current block is in general surrounded by partially reconstructed image signals, and therefore, the interpolative prediction can be performed in almost all the blocks. In addition, it requires no extra side information as far as the coding order of DCT coefficients is fixed in a whole image.

The rest of the paper describes the details of the proposed intra prediction method. Simulation results are also shown to confirm its effectiveness in practical DCT-based image coding.

2. COEFFICIENT-WISE INTRA PREDICTION FOR DCT COEFFICIENTS

In this section, we assume that sets of DCT coefficients, which are obtained from an entire image composed of M blocks, are encoded in coefficient-wise order from lower to higher spatial frequencies. Specifically, two dimensional DCT coefficients in a block of size $N \times N$ are rearranged in zigzag scan order as shown in Fig. 1 and the coefficients with the same frequency index i are grouped and encoded at the i-th pass ($i = 0, 1, 2, \ldots, N^2 - 1$), while the blocks are visited in raster scan order ($b = 0, 1, 2, \ldots, M-1$) at every coding pass. Such a coding process realizes scalable representation of image contents and, for instance, is adopted in spectral selection coding for the progressive JPEG [9].

Let us consider that the DCT coefficient $y_{b,i}$ which belongs to the b-th block is being encoded at the i-th pass. Already encoded DCT coefficients are stored in memory and a frame buffer containing the reconstructed image is updated at every encoding stage of a single coefficient. In this case, the current block is in general surrounded by partially reconstructed image signals as shown in Fig. 2. Four blocks of top-left, top, top-right and left neighbors contain image signals $f_i(x, y)$ reconstructed from $i+1$ DCT coefficients at each block, i.e. DCT coefficients with frequency indices higher than i are set to zero in the reconstruction process using inverse DCT. Quality of the reconstructed signals $f_{i-1}(x, y)$ in the remaining four neighboring blocks (right, bottom-left, bottom and bottom-right) is slightly worse because DCT coefficients with the frequency index i are unavailable in those blocks. In either case, the reconstructed signals at the

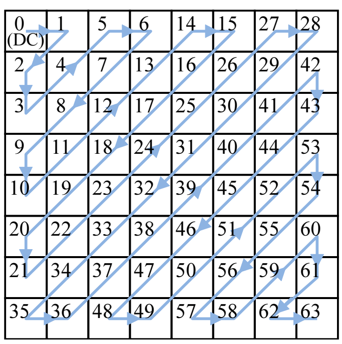

Fig. 1. Zigzag scanning of DCT coefficients ($N = 8$).

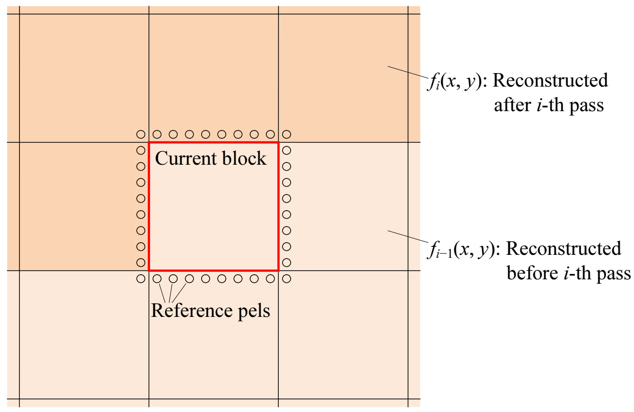

Fig. 2. Reference pels used in the proposed intra prediction.

reference pels at any sides of the current block can be used for the prediction. In the proposed method, $4N+4$ reference pels are disposed around the current block as shown by small circles in Fig. 2. Furthermore, prediction modes used in the H.265/HEVC standard are modified to enable interpolative prediction which utilize these reference pels.

2.1. Interpolative planar prediction

Planar prediction used in the proposed method generates smooth values in a block by weighting reference pels taken from four directions. In Fig. 3, a pel $\boldsymbol{q}$ is being predicted and $\boldsymbol{p}_N, \boldsymbol{p}_W, \boldsymbol{p}_E$ and $\boldsymbol{p}_S$ are reference pels which share the same horizontal or vertical coordinate as $\boldsymbol{q}$. A predicted value at the pel $\boldsymbol{q}$ is calculated by:

$$\hat{f}_i(\boldsymbol{q}) = \frac{1}{2N+2}\Big\{(N-x)f_i(\boldsymbol{p}_W) + (x+1)f_{i-1}(\boldsymbol{p}_E) + (N-y)f_i(\boldsymbol{p}_N) + (y+1)f_{i-1}(\boldsymbol{p}_S)\Big\}, \quad (1)$$

where x and y indicate the position of the pel $\boldsymbol{q}$ in a local coordinate whose origin is placed at the most upper left pel

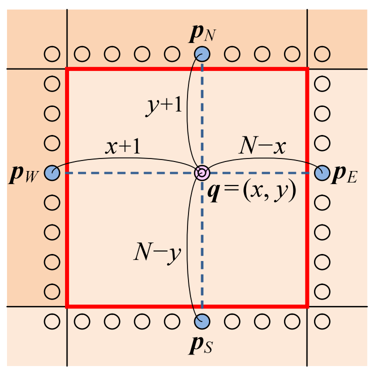

Fig. 3. Interpolative planar prediction.

in the current block. In the H.265/HEVC standard [5], the reconstructed values at the reference pels $\boldsymbol{p}_E$ and $\boldsymbol{p}_S$ are unavailable and their substitutes are taken from top-right and bottom-left corners of the block boundary, respectively.

2.2. Interpolative DC prediction

In DC prediction, the reconstructed values at the reference pels shown as blue circles in Fig. 4 are once averaged:

$$\bar{f}_i = \frac{1}{4N+4}\left\{\sum_{\boldsymbol{p}\in\boldsymbol{R}_A} f_i(\boldsymbol{p}) + \sum_{\boldsymbol{p}\in\boldsymbol{R}_B} f_{i-1}(\boldsymbol{p})\right\}, \tag{2}$$

where the reference pels are categorized into two sets $\boldsymbol{R}_A$ and $\boldsymbol{R}_B$ according to whether the i-th coding pass is already done or not. We include the reference pels at right and bottom sides ($\boldsymbol{R}_B$) in this averaging, whereas they are unavailable in the H.265/HEVC case. Then the current block is filled with the averaged value $\bar{f}_i$ and the border pels shown as dotted circles in Fig. 4 are filtered in the same way as the H.265/HEVC standard [5]. As a result, predicted values at the block border are smoothly connected to the surrounding blocks. In general, AC components of the predicted values are produced through this filtering process.

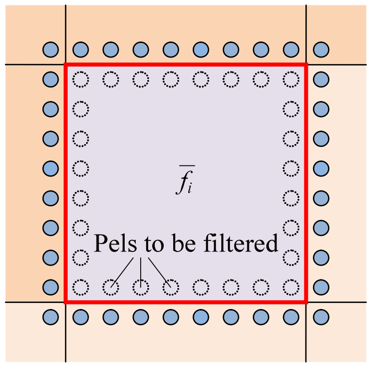

Fig. 4. Interpolative DC prediction.

2.3. Directional prediction

Similarly to the angular prediction [5], directional prediction used in the proposed method is defined on the designated prediction direction. In Fig. 5, a straight line L is passing through the pel $\boldsymbol{q}$ to be predicted and its slope (angle of inclination) is given by θ as the prediction direction. The line L has two cross points $\boldsymbol{p}_F$ and $\boldsymbol{p}_B$ with a square which is formed by connecting adjacent reference pels. Since the pel $\boldsymbol{q}$ lies between $\boldsymbol{p}_F$ and $\boldsymbol{p}_B$, it is reasonable to use both the cross points to obtain the predicted value $\hat{f}_i(\boldsymbol{q})$. In [10], a similar technique, which blends two reference samples using a Gaussian-based weight function, is applied only when the samples are available on both sides. In our method, both of the samples (cross points) are always available around the block boundary and the predicted value at the pel $\boldsymbol{q}$ is linearly interpolated:

$$\hat{f}_i(\boldsymbol{q}) = \frac{||\boldsymbol{p}_B-\boldsymbol{q}||\tilde{f}_i(\boldsymbol{p}_F) + ||\boldsymbol{p}_F-\boldsymbol{q}||\tilde{f}_i(\boldsymbol{p}_B)}{||\boldsymbol{p}_F-\boldsymbol{p}_B||}. \tag{3}$$

Values at the cross points $\tilde{f}_i(\boldsymbol{p}_F)$ and $\tilde{f}_i(\boldsymbol{p}_B)$ are also calculated as linear interpolation of adjacent reference pels. In this case, a reconstructed value at the reference pel $\boldsymbol{p}$ is defined as:

$$\tilde{f}_i(\boldsymbol{p}) = \begin{cases} f_i(\boldsymbol{p}) & (\boldsymbol{p}\in\boldsymbol{R}_A) \\ f_{i-1}(\boldsymbol{p}) & (\text{otherwise}) \end{cases}. \tag{4}$$

By changing the prediction angle θ with a step of $\pi/32$, we prepare 32 kinds of the directional prediction modes in the experiments shown later.

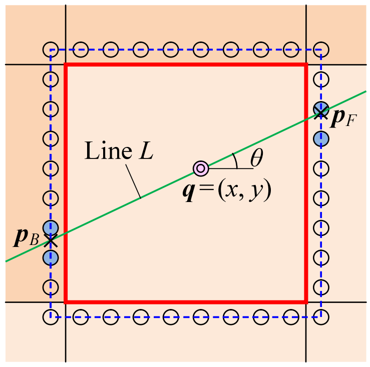

Fig. 5. Directional prediction.

2.4. Calculation of prediction residuals

In the proposed method, the predicted image is separately generated at each coding pass. However, the prediction mode is common for all the coding pass in the same block to reduce side information needed for the adaptive intra prediction. A predicted value $\hat{y}_{b,i}$ is obtained by applying DCT to the reconstructed image and a prediction residual $e_{b,i}$ is calculated in DCT domain:

$$e_{b,i} = y_{b,i} - \hat{y}_{b,i}, \tag{5}$$

where $y_{b,i}$ is a DCT coefficient of the original image to be encoded in the b-th block at the i-th coding pass. In general, the value of $e_{b,i}$ is immediately quantized and entropy coded to form the compressed bitstream.

3. HYBRID APPROACH OF BLOCK-WISE AND COEFFICIENT-WISE INTRA PREDICTION

The above mentioned coefficient-wise intra prediction method can be accurate when the reference pels used for the interpolative prediction have rich information on the image signals. Unfortunately, this is not the case at an early stage of the coding passes because the reconstructed values at the reference pels are reproduced from only low frequency components of DCT coefficients. To cope with this problem, we divide the DCT coefficients into low and high frequency classes according to a given threshold T and apply the coefficient-wise intra prediction only to the high frequency class $\{y_{b,i} \,|\, i \geq T\}$. Preparatory to this, the block-wise intra prediction which deals together with T DCT coefficients in each block is carried out for the low frequency class $\{y_{b,i} \,|\, i < T\}$ in a similar way to the conventional intra prediction. In our implementation, planar and DC prediction modes are identical to those in the H.265/HEVC standard. On the other hand, the directional prediction modes are borrowed from [8] to allow interpolative prediction as far as possible. In the case of $T = 64$, only the block-wise intra prediction is conducted in a single coding pass.

4. SIMULATION RESULTS

In order to evaluate effectiveness of the coefficient-wise intra prediction, we have implemented the proposed method in the actual coding scheme which was developed by the authors for lossless re-encoding of JPEG images [11]. For further compression of the existing JPEG images, the scheme exploited inter-block correlations of DCT coefficients extracted from a JPEG bitstream by applying the H.264/AVC-based intra prediction [4]. In this case, the DCT coefficient $y_{b,i}$ in (5) is already quantized and the prediction residual $e_{b,i}$ is still discrete. This allows lossless encoding of the prediction residual without any quantization process. In addition, bypassing the quantization process offers an advantage in that the best prediction mode can be selected without causing interference with other blocks since the reconstructed values used in the prediction can be treated as constant. In the present implementation, the situation is same but the intra prediction is just replaced by the proposed method.

Fig. 6 reports bitrate savings achieved by the proposed method. In this experiment, two types of JPEG images encoded with different quantization tables are tested. 'Standard Q table' is based on annex K of the JPEG standard [1] (we use the IJG software [12] with a quality factor of 75) while 'flat Q table' imposes the same quantization step size ($\Delta Q = 20$) for all the frequency components. We can see that the 'flat Q table' generally provides larger gains than the 'standard Q table' when the threshold for the frequency index i is $T > 32$. This is because high frequency components of DCT coefficients, for which the coefficient-wise intra prediction seems to be suitable, are suppressed utilizing insensitivity of the human visual system in the case of 'standard Q table'. For the same reason, the coding gains for 'Balloon' which originally contains less high frequency components are almost spoiled. As expected in section 3, use of smaller values for the threshold T deteriorates the coding performance due to poor prediction results for low frequency components.

Fig. 7 plots rate-PSNR curves of our coding scheme to-

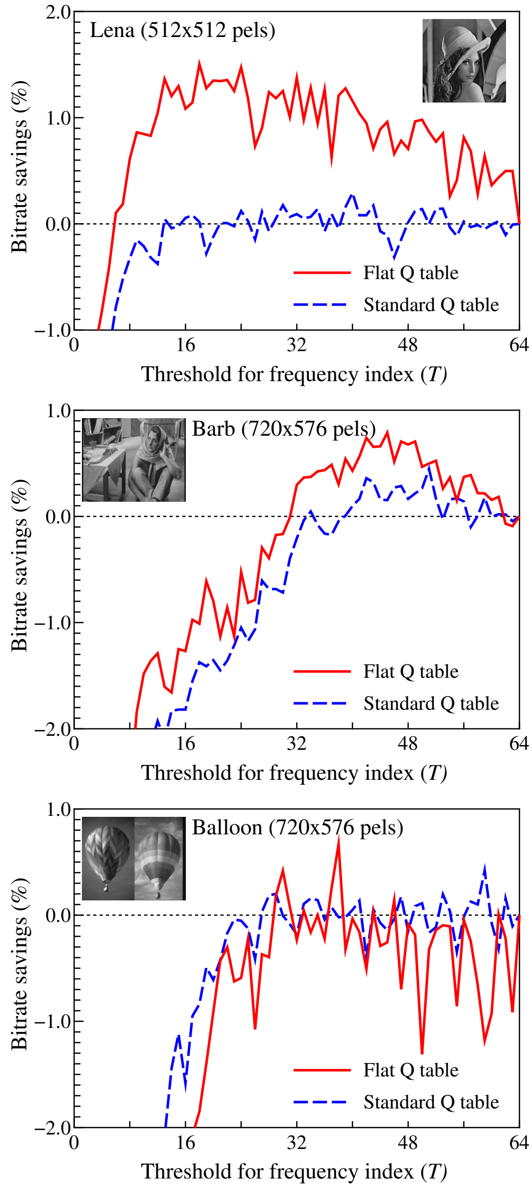

Fig. 6. Bitrate savings over the block-wise intra prediction ($T = 64$).

gether with those of the standard coding schemes: JPEG [1], JPEG 2000 (9/7 wavelet) [13] and H/265/HEVC intra-frame coding [5]. In this figure, texts highlighted in red mean our lossless re-encoding scheme with different intra prediction techniques. It is confirmed that the coding efficiency of our scheme with the proposed intra prediction method ($T = 36$) is comparable to that of the JPEG 2000 standard. Though the latest H.265/HEVC standard significantly outperforms the other schemes mainly due to variable block sizes in prediction and transform processes, our scheme has a unique feature of lossless re-encoding, that is, the bitstream can be transcoded into JPEG compliant one without any loss of quality [11]. In Fig. 8, close-up views of the predicted images generated from

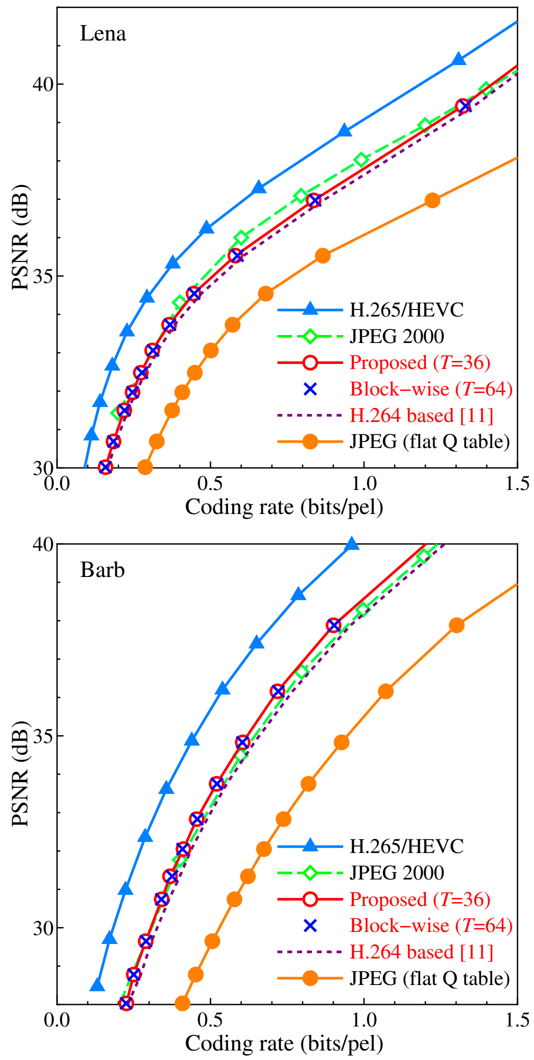

Fig. 7. Coding performance.

(a) Proposed ($T = 36$) (b) Block-wise ($T = 64$)

Fig. 8. Close-up views of the predicted images (Barb).

the predicted values of the DCT coefficients $\{\hat{f}_{b,i}\}$ are shown as an example ($\Delta Q = 20$). A striped pattern of the scarf around woman's head is well reproduced when the hybrid approach of coefficient-wise and block-wise intra prediction ($T = 36$) is employed

5. CONCLUSIONS

In this paper, a new intra prediction method which is designed to exploit inter-block correlation for DCT based image coding is proposed. By changing coding process of DCT coefficients from block-wise order to coefficient-wise one, the method can perform interpolative prediction in almost all blocks. Simulation results indicate that a hybrid approach of block-wise and coefficient-wise orders provides better coding performance.

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] G. Cote, B. Erol, M. Gallant and F. Kossentini, "H.263+: Video Coding at Low Bit Rates," IEEE Trans. on Circuits and Syst. for Video Technol., vol.8, no.7, pp.849–866, Nov. 1998.

[3] T. Ebrahimi and C. Horne, "MPEG-4 Natural Video Coding – An Overview," Signal Processing: Image Communication, vol.15, no.4, pp.365–385, Jan. 2000.

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

[5] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang and T. Wiegand, "High Efficiency Video Coding (HEVC) Text Specification Draft 10 (for FDIS & Last Call)," JCT-VC, Doc. JCTVC-L1003, Jan. 2013.

[6] B. Girod, "Why B-Pictures Work: A Theory of Multi-Hypothesis Motion-Compensated Prediction," Proc. of IEEE Intl. Conf. on Image Processing (ICIP '98), Vol.II, pp.213–217, Oct. 1998.

[7] T. Shiodera, A. Tanizawa and T. Chujoh, "Block Based Extra/Inter-Polating Prediction for Intra Coding," Proc. of IEEE Intl. Conf. on Image Processing (ICIP 2007), Vol.6, pp.445–448, Sep. 2007.

[8] I. Matsuda, Y. Ohtake, S. Mochizuki, H. Fukai and S. Itoh, "Interpolative Intra Prediction by Adapting Processing Order in Block-Based Image Coding," Proc. of IEEE Intl. Conf. on Image Processing (ICIP 2013), pp.1646–1650, Sep. 2013.

[9] J. In, S. Shirani and F. Kossentini, "JPEG Compliant Efficient Progressive Image Coding," Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP '98), Vol.5, pp.2633–2636, May 1998.

[10] S. Matsuo, S. Takamura, H. Fujii and A Shimizu, "AHG7: Modification of Intra Angular Prediction Blending," ITU-T VCEG and ISO/IEC MPEG, Doc. JCTVC-L0128, Jan. 2013.

[11] 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), L3-6, Aug. 2008.

[12] Independent JPEG Group: http://www.ijg.org/

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