
ELSEVIER

Pattern Recognition Letters 22 (2001) 1337–1345

Pattern Recognition Letters

www.elsevier.com/locate/patrec

JPEG2000: The upcoming still image compression standard

A.N. Skodras [a,*], C.A. Christopoulos [b], T. Ebrahimi [c]

[a] *Electronics Laboratory, University of Patras, GR-26110 Patras, Greece*
[b] *Media Lab, Ericsson Research, Ericsson Radio Systems AB, S-16480 Stockholm, Sweden*
[c] *Signal Processing Laboratory, EPFL, CH-1015 Lausanne, Switzerland*

Abstract

With the increasing use of multimedia technologies, image compression requires higher performance as well as new features. To address this need in the specific area of still image encoding, a new standard is currently being developed, the JPEG2000. It is not only intended to provide rate-distortion and subjective image quality performance superior to existing standards, but also to provide functionality that current standards can either not address efficiently or not address at all.

Keywords: JPEG; Colour image compression; Source coding; Subband coding; Wavelet transform

1. Introduction

Since the mid-1980s, members from the International Telecommunication Union (ITU) and the International Organisation for Standardisation (ISO) have been working together to establish a joint international standard for the compression of continuous-tone (multilevel) still images, both greyscale and colour. This effort has been known as JPEG, the Joint Photographic Experts Group. (The 'joint' in JPEG refers to the collaboration between ITU and ISO). Officially, JPEG corresponds to the ISO/IEC international standard 10928-1, digital compression and coding of continuous-tone still images or to the ITU-T Recommendation T.81. The text in both these ISO and ITU-T documents is identical. JPEG became a draft international standard (DIS) in 1991 and an international standard (IS) in 1992 (Pennebaker and Mitchell, 1993).

With the continual expansion of multimedia and Internet applications, the needs and requirements of the technologies used, grew and evolved. In March 1997 a new call for contributions was launched for the development of a new standard for the compression of still images, the JPEG2000. This project, JTC [1] 1.29.14 (15444), was intended to create a new image coding system for different types of still images (bi-level, grey-level, colour, multi-component), with different characteristics (natural images, scientific, medical, remote sensing, text, rendered graphics, etc.) allowing different imaging models (client/server, real-time transmission, image library archival, limited buffer and bandwidth resources, etc.) preferably within a

* Corresponding author. Fax: +30-61-997-456.
E-mail addresses: skodras@cti.gr (A.N. Skodras), charilaos.christopoulos@era.ericsson.se (C.A. Christopoulos), Touradj.Ebrahimi@epfl.ch (T. Ebrahimi).

[1] JTC stands for Joint Technical Committee.

0167-8655/01/$ - see front matter
PII: S0167-8655(01)00079-4

unified system. This coding system should provide low bit-rate operation with rate-distortion and subjective image quality performance superior to existing standards, without sacrificing performance at other points in the rate-distortion spectrum, incorporating at the same time many contemporary features. The standardisation process, which is co-ordinated by the JTC1/SC29/WG1 of ISO/IEC [2] has already (as of May 2000) produced the Final Committee Draft (FCD) of the JPEG2000 part I decoder (ISO/IEC, 2000). The IS is scheduled for December 2000.

The JPEG2000 standard provides a set of features that are of vital importance to many high-end and emerging applications, by taking advantage of new technologies. It addresses areas where current standards fail to produce the best quality or performance and provides capabilities to markets that currently do not use compression. The markets and applications better served by the JPEG2000 standard are Internet, colour facsimile, printing, scanning (consumer and pre-press), digital photography, remote sensing, mobile, medical imagery, digital libraries/archives and E-commerce. Each application area imposes some requirements that the standard should fulfil (Requirements AHG, 1999). The main features that this standard possesses are: superior low bit-rate performance, continuous-tone and bi-level compression, lossless and lossy compression, progressive transmission by pixel accuracy and resolution, random codestream access and processing, robustness to bit-errors.

In this paper the structure of the JPEG2000 standard is presented and performance comparisons are reported. The paper is organised in the following way: In Section 2 the architecture of the standard is described, and in Section 3 the multiple-component case is covered. The file format aspects and other interesting features of the standard, like region-of-interest coding, error resilience and scalability are presented in Section 4. Finally, some performance comparisons are reported in Section 5 of the paper.

[2] SC, WG, IEC stand for Standing Committee, Working Group and International Electrotechnical Commission, respectively.

2. Architecture of the standard

The block diagram of the JPEG2000 encoder is illustrated in Fig. 1(a). A discrete wavelet transform (DWT) is first applied on the source image data. The transform coefficients are then quantised and entropy coded, before forming the output codestream (bit-stream). At the decoder (Fig. 1(b)), the codestream is first entropy decoded, dequantised and inverse discrete transformed, providing the reconstructed image data. It is worth mentioning that, unlike other coding schemes, the JPEG2000 can be both lossy and lossless. This depends on the wavelet transform and the quantisation applied.

Before proceeding with the details of each block of Fig. 1, it should be mentioned that the standard works on image tiles. The term 'tiling' refers to the partition of the original (source) image into rectangular non-overlapping blocks (tiles), which are compressed independently, as though they were entirely distinct images (Fig. 2). This is the strongest form of spatial partitioning, in that all operations, including component mixing, wavelet transform, quantisation and entropy coding are performed independently on the different tiles of the image. All tiles have exactly the same dimensions, except maybe those, which abut the right and lower boundary of the image. Arbitrary tile sizes are allowed, up to and including the entire image (i.e. no tiles). Tiling reduces memory requirements and constitutes one of the methods for the efficient extraction of a region of the image. Prior to computation of the forward DWT on each tile, all samples of the image tile component are DC level shifted by subtracting the same quantity (i.e. the component depth) from each sample (Fig. 2).

2.1. The wavelet transform

Tile components are decomposed into different decomposition levels using a wavelet transform. These decomposition levels contain a number of subbands populated with coefficients that describe

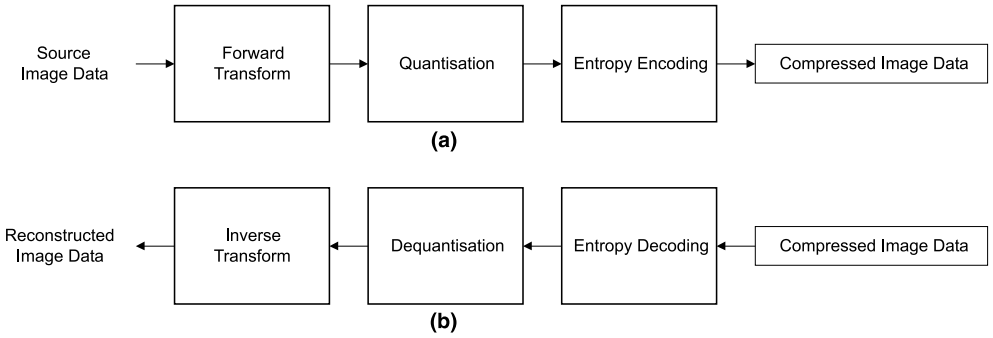

Fig. 1. Block diagram of the JPEG2000 (a) encoder and (b) decoder.

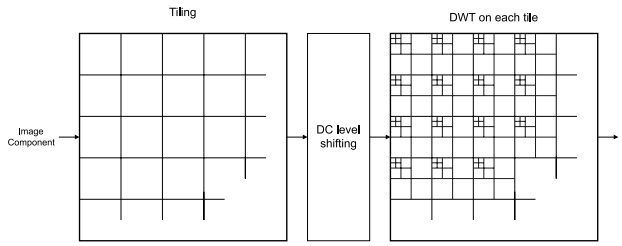

Fig. 2. Tiling, DC level shifting and DWT of each image tile component.

the horizontal and vertical spatial frequency characteristics of the original tile component planes (Fig. 2). The coefficients provide local frequency information. A decomposition level is related to the next decomposition level by spatial powers of two. Part I of the standard supports dyadic decomposition, since this appears to yield the best compression performance for natural images. To perform the forward DWT the standard uses a 1D subband decomposition of a 1D set of samples into low-pass samples, representing a downsampled low resolution version of the original set, and high-pass samples, representing a downsampled residual version of the original set, needed for the perfect reconstruction of the original set from the low-pass set. In general, any user supplied wavelet filter bank may be used. The DWT can be *irreversible* or *reversible*. The default (Part I) *irreversible transform* is implemented by means of the Daubechies 9-tap/7-tap filter (Antonini et al., 1992). The analysis and the corresponding synthesis filter coefficients are given in Table 1. The default (Part I) *reversible transformation* is implemented by means of the 5-tap/3-tap filter, the coefficients of which are given in Table 2 (Le Gall and Tabatabai, 1988).

Table 1
Daubechies 9/7 analysis and synthesis filter coefficients

i	Analysis filter coefficients		Synthesis filter coefficients	
	Low-pass filter $h_L(i)$	High-pass filter $h_H(i)$	Low-pass filter $g_L(i)$	High-pass filter $g_H(i)$
0	0.6029490182363579	1.115087052456994	1.115087052456994	0.6029490182363579
±1	0.2668641184428723	−0.5912717631142470	0.5912717631142470	−0.2668641184428723
±2	−0.07822326652898785	−0.05754352622849957	−0.05754352622849957	−0.07822326652898785
±3	−0.01686411844287495	0.09127176311424948	−0.09127176311424948	0.01686411844287495
±4	0.02674875741080976			0.02674875741080976

Table 2
5/3 Analysis and synthesis filter coefficients

i	Analysis filter coefficients		Synthesis filter coefficients	
	Low-pass filter $h_L(i)$	High-pass filter $h_H(i)$	Low-pass filter $g_L(i)$	High-pass filter $g_H(i)$
0	6/8	1	1	6/8
±1	2/8	−1/2	1/2	−2/8
±2	−1/8			−1/8

The standard supports two filtering modes: a *convolution-based* and a *lifting-based*. For both modes to be implemented, the signal should be first extended periodically as shown in Fig. 3. This *periodic symmetric extension* is used to ensure that for the filtering operations that take place at both boundaries of the signal, one signal sample exists and spatially corresponds to each coefficient of the filter mask. The number of additional samples required at the boundaries of the signal is therefore filter-length dependent (ISO/IEC, 2000).

Convolution-based filtering consists in performing a series of dot products between the two filter masks and the extended 1D signal. *Lifting-based filtering* consists of a sequence of very simple filtering operations for which alternately odd sample values of the signal are updated with a weighted sum of even sample values, and even sample values are updated with a weighted sum of odd sample values (ISO/IEC, 2000; Calderbank et al., 1997; Kovacevic and Sweldens, 2000). For the reversible (lossless) case the results are rounded to integer values. The lifting-based filtering for the 5/3 analysis filter is achieved by means of Eq. (1)

$$\begin{aligned} y(2n+1) &= x_{\text{ext}}(2n+1) \\ &\quad - \left\lfloor \frac{1}{2}[x_{\text{ext}}(2n) + x_{\text{ext}}(2n+2)] \right\rfloor, \\ y(2n) &= x_{\text{ext}}(2n) \\ &\quad + \left\lfloor \frac{1}{4}(y(2n-1) + y(2n+1)) + \frac{1}{2} \right\rfloor, \end{aligned} \tag{1}$$

Fig. 3. Periodic symmetric extension of a signal.

where x_{ext} is the extended input signal, y is the output signal and $\lfloor a \rfloor$ indicates the largest integer not exceeding a.

2.2. Quantisation

Quantisation is the process by which the transform coefficients are reduced in precision. This operation is lossy, unless the quantisation step is 1 and the coefficients are integers, as produced by the reversible integer 5/3 wavelet. Each of the transform coefficients $a_b(u,v)$ of the subband b is quantised to the value $q_b(u,v)$ (i.e. scalar quantisation) according to the formula

$$q_b(u,v) = \text{sign}(a_b(u,v)) \left\lfloor \frac{|a_b(u,v)|}{\Delta_b} \right\rfloor, \tag{2}$$

where Δ_b is the quantisation step of subband b. One quantisation step per subband is allowed. All quantised transform coefficients are signed values even when the original components are unsigned. These coefficients are expressed in a sign–magnitude representation prior to coding.

2.3. Entropy coding

Each subband of the wavelet decomposition is divided up into rectangular blocks, called *code-blocks*, which are coded independently using arithmetic coding. This approach, called EBCOT (embedded block coding with optimised truncation), was introduced in 1998 (Taubman, 1998, 2000). Such a partitioning reduces memory requirements in both hardware and software implementations and provides a certain degree of spatial random access to the bit-stream. The block size is identical for all subbands, so that blocks in lower resolution subbands span a larger

region in the original image. A neighbourhood of spatially consistent code-blocks from each subband at a given resolution level forms larger rectangles, called *precincts*. Code-blocks are coded at a bit-plane at a time, starting with the most significant bit-plane with a non-zero element to the least significant bit-plane. For each bit-plane in a code-block, a special code-block scan pattern is used for each of the three passes, i.e. the *significance propagation pass*, the *magnitude refinement pass* and the *clean-up pass* (Marcellin et al., 2000). Each coefficient bit in the bit-plane is coded in only one of the three passes. A rate distortion optimisation method is used to allocate a certain number of bits to each block. The recursive probability interval subdivision of Elias coding is the basis for the binary *arithmetic coding* process. With each binary decision, the current probability interval is subdivided into two subintervals, and the codestream is modified (if necessary) so that it points to the base (the lower bound) of the probability subinterval assigned to the symbol which occurred. Since the coding process involves addition of binary fractions rather than concatenation of integer codewords, the more probable binary decisions can often be coded at a cost of much less than one bit per decision (ISO/IEC, 2000).

3. Multiple-component images

JPEG2000 supports multiple-component images. Different components need not have the same bit-depths; nor need they have all been signed or unsigned. For reversible systems, the only requirement is that the bit-depth of each output image component must be identical to the bit-depth of the corresponding input image component.

The standard supports two different component transformations: one *irreversible component transformation* (ICT) and one *reversible component transformation* (RCT). The block diagram of the JPEG2000 colour image (RGB) encoder is shown in Fig. 4. C_1, C_2, C_3 represent in general the colour transformed output components. If needed, prior to applying the forward colour transformation, the image component samples are DC level shifted.

The ICT may only be used for lossy coding. It can be seen as an approximation of a YC_bC_r transformation of the RGB components. The forward and the inverse irreversible component transformations are already known well (ISO/IEC, 2000; Pennebaker and Mitchell, 1993).

The RCT may be used for lossy or lossless coding. It is a decorrelating transformation, which is applied to the three first components of an image. Three goals are achieved by this transformation, namely, colour decorrelation for efficient compression, reasonable colour space with respect to the Human Visual System for quantisation, and ability of having lossless compression, i.e. exact reconstruction with finite integer precision. For the RGB components, the RCT can be seen as an approximation of a YUV transformation. The forward and inverse RCT is performed by means of Eq. (3)

$$\begin{aligned} Y_r &= \lfloor (R + 2G + B)/4 \rfloor, & G &= Y_r - \lfloor (U_r + V_r)/4 \rfloor, \\ U_r &= B - G, & R &= V_r + G, \\ V_r &= R - G, & B &= U_r + G. \end{aligned} \tag{3}$$

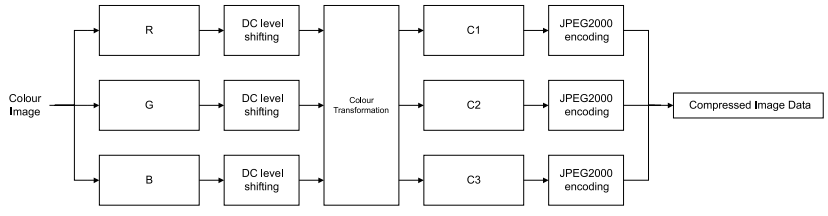

Fig. 4. Block diagram of the JPEG2000 colour image encoder.

4. Significant features of the standard

The JPEG2000 standard exhibits a lot of features, the most significant being the possibility to define regions of interest in an image, the spatial and SNR scalability, the error resilience and the possibility of intellectual property rights protection. Interestingly enough, all these features are incorporated within a unified algorithm.

4.1. Region-of-interest (ROI)

One of the features included in JPEG2000 is the ROI coding. According to this, certain ROIs of the image can be coded with better quality than the rest of the image (background). The ROI *scaling-based* method used scales up (DC shifts) the coefficients so that the bits associated with the ROI are placed in higher bit-planes. During the embedded coding process, these bits are placed in the bit-stream before the non-ROI parts of the image. Thus, the ROI is decoded or refined, before the rest of the image. Regardless of the scaling, a full decoding of the bit-stream results in a reconstruction of the whole image with the highest fidelity available. If the bit-stream is truncated, or the encoding process is terminated before the whole image is fully encoded, the ROI is of higher fidelity than the rest of the image. The ROI approach defined in the JPEG2000 Part I is called MAXSHIFT method and allows ROI encoding of arbitrary shaped regions without the need of shape information and shape decoding (Christopoulos et al., 2000).

4.2. Scalability

Realising that many applications require images to be simultaneously available for decoding at a variety of resolutions or qualities, the JPEG2000 architecture supports scalability. In general, scalable coding of still images means the ability to achieve coding of more than one resolution and/or quality simultaneously. Scalable image coding involves generating a coded representation (bit-stream) in a manner which facilitates the derivation of images of more than one resolution and/or quality by scalable decoding. Bit-stream scalability is the property of a bit-stream that allows decoding of appropriate subsets of the bit-stream to generate complete pictures of resolution and/or quality commensurate with the proportion of the bit-stream decoded. For scalable bit-streams, decoders of different complexities, from low performance to high performance, can coexist. While low performance decoders may decode only small portions of the bit-stream producing basic quality, high performance decoders may decode much more and produce significantly higher quality. The most important types of scalability are signal-to-noise ratio (SNR) scalability and spatial scalability (ISO/IEC, 2000; Marcellin et al., 2000). *SNR scalability* involves generating at least two image layers of same spatial resolution, but different qualities, from a single image source. The lower layer is coded by itself to provide the basic image quality and the enhancement layers are coded to enhance the lower layer. The enhancement layer, when added back to the lower layer, regenerates a higher quality reproduction of the input image. *Spatial scalability* involves generating at least two spatial resolution layers from a single source such that the lower layer is coded by itself to provide the basic spatial resolution and the enhancement layer employs the spatially interpolated lower layer and carries the full spatial resolution of the input image source. An additional advantage of spatial and SNR scalability types is their ability to provide resilience to transmission errors, as the most important data of the lower layer can be sent over a channel with better error performance, while the less critical enhancement layer data can be sent with poor error performance. Both types of scalability are very important for Internet and database access applications and bandwidth scaling for robust delivery. The SNR and spatial scalability types include the progressive and hierarchical coding modes already defined in the current JPEG, but they are more general.

4.3. Error resilience

Many applications require the delivery of image data over different types of communication channels. Typical wireless communication channels give rise to random and burst bit-errors. Internet

communications are prone to loss due to traffic congestion. To improve the performance of transmitting compressed images over these error prone channels, error resilient bit-stream syntax and tools are included in the JPEG2000 standard. The error resilience tools deal with channel errors using approaches like data partitioning and re-synchronisation, error detection and concealment, and quality of service (QoS) transmission based on priority (ISO/IEC, 2000).

4.4. New file format with IPR capabilities

An optional file format (JP2) for the JPEG2000 compressed image data has been defined in the standard. This format has got provisions for both image and metadata, a mechanism to indicate the tonescale or colourspace of the image, a mechanism by which readers may recognise the existence of intellectual property rights (IPR) information in the file and a mechanism by which metadata (including vendor specific information) can be included in the file (ISO/IEC, 2000).

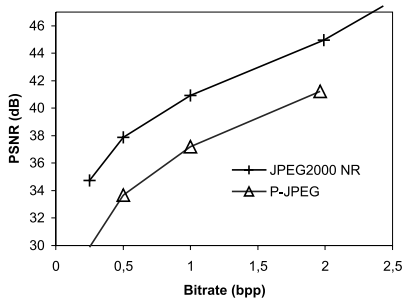

Fig. 5. Rate-distortion results for the JPEG2000 versus the progressive JPEG for a natural image.

5. Comparative results

The rate-distortion behaviour of the lossy (non-reversible) JPEG2000 and the progressive JPEG is depicted in Fig. 5 for a natural image. It is seen that the JPEG2000 significantly outperforms the JPEG scheme for any given rate. We can easily conclude that for similar PSNR quality, the JPEG2000 compresses almost twice more than JPEG (Charrier et al., 1999; Christopoulos and Skodras, 1999). The superiority of the JPEG2000 can be subjectively judged with the help of Fig. 6, where the reconstructed image 'hotel' (720×576) is shown. This image was compressed at a rate of 0.125 bpp using the existing JPEG and the upcoming JPEG2000 (http://etro.vub.ac.be/~chchrist/jpeg2000_contributions.htm). From the point of view of visual quality, JPEG2000 is 10–25% better than the baseline JPEG (for images compressed at approximately 0.5–1 bpp). The improvement is much higher in the case of very low bit-rates.

One of the interesting and unique features of JPEG2000 is its capability in defining ROIs, that are coded at a better quality than the rest of the

Fig. 6. Reconstructed image 'hotel' compressed at 0.125 bpp by means of (a) JPEG and (b) JPEG2000.

image. ROIs can be more than one and of any shape and size. In Fig. 7 an example of a circular ROI is given. Experiments have shown that for lossless coding of images, the ROI feature results in an increase of the bit-rate by a maximum of 8% in comparison to lossless coding without using the ROI feature (Christopoulos et al., 2000).

The lossless compression efficiency of JPEG2000 versus the lossless mode of JPEG and JPEG-LS for a natural and a compound image is reported in Table 3. It is seen that JPEG2000 performs equivalently to JPEG-LS in the case of the natural image, with the added benefit of scalability. JPEG-LS, however, is advantageous in the case of the compound image. Taking into account that JPEG-LS is significantly less complex than JPEG2000, it is reasonable to use JPEG-LS for lossless compression. In such a case though, the generality of JPEG2000 is sacrificed. A comparison of JPEG, JPEG-LS and JPEG2000 from the functionality point of view is illustrated in Table 4. A plus (or minus) sign indicates that the corresponding functionality is supported (or not supported). The more the plus signs the greater the support. The parentheses indicate that a separate mode is required. It becomes evident from Table 4, that the JPEG2000 standard offers the richest set of features in a very efficient way and within a unified algorithm.

Fig. 7. Reconstructed image in which an ROI of circular shape has been defined.

But, all of the above-mentioned advantages of JPEG2000 are at the expense of memory and computational complexity. Optimised JPEG codecs run almost three times faster and require less memory than current JPEG2000 software implementations. It should be stressed though, that these figures refer to non-optimal implementations, which are also platform dependent. Careful optimisation of the JPEG2000 algorithm will greatly improve performance without sacrificing functionality. However, the multi-pass bit-plane context model and the arithmetic coder of the JPEG2000 will prevent any software implementation from reaching the speed JPEG obtains with DCT and Huffman coder. As for the memory requirements, the new line-based reduced memory approaches for the calculation of the wavelet

Table 3
Lossless compression results (in bpp)

Image	Lossless JPEG	JPEG-LS	JPEG2000
Lena (512×512, 24 bpp)	14.75	13.56	13.54
Cmpnd1 (512×768, 8 bpp)	2.48	1.24	2.12

Table 4
Summary of functionalities supported by each standard

Compression algorithm	Lossless	Lossy	Embedded bit-stream	ROI	Error resilience	Scalability	Complexity	Random access	Generic
JPEG	(+)	++	–	–	–	(+)	+(+)	+	+
JPEG-LS	++++	+	+	–	–	–	+	–	+
JPEG2000	+++	+++	+++	++	++	++	+++	++	+++

transform offer already less memory hungry implementations, as compared to the JPEG ones, especially in the progressive coding case (Chrysafis and Ortega, 2000).

6. Conclusions

JPEG2000 is the new standard for still image compression that is going to be in use by the beginning of next year. It provides a new framework and an integrated toolbox to better address increasing needs for compression and functionalities for still image applications, like Internet, colour facsimile, printing, scanning, digital photography, remote sensing, mobile applications, medical imagery, digital library and E-commerce. Lossless and lossy coding, embedded lossy to lossless, progressive by resolution and quality, high compression efficiency, error resilience and lossless colour transformations are some of its features. Comparative results have shown that JPEG2000 is indeed superior to existing still image compression standards. Work is still needed in optimising its implementation performance. The reference software of the standard has been developed in C and in JAVA. The intention is to have a license fee-free software for commercial and non-commercial use. The JAVA version can be downloaded from http://jj2000.epfl.ch.

Acknowledgements

The authors would like to thank Diego Santa Cruz from EPFL and Mathias Larsson and Joel Askelof from Ericsson, Sweden for their contribution in this work.

References

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

Calderbank, A.R., Daubechies, I., Sweldens, W., Yeo, B.-L., 1997. Lossless image compression using integer to integer wavelet transforms. In: Proc. Internat. Conf. on Image Processing.

Charrier, M., Santa Cruz, D., Larsson, M., 1999. JPEG2000: The next millennium compression standard for still images. In: Proc. ICMCS-99.

Christopoulos, C., Askelof, J., Larsson, M., 2000. Efficient methods for encoding regions of interest in the upcoming JPEG2000 still image coding standard. IEEE Signal Process. Lett. 7 (9), 247–249.

Christopoulos, C.A., Skodras, A.N., 1999. The JPEG2000 (Tutorial). In: Proc. IEEE Internat. Conf. on Image Processing.

Chrysafis, C., Ortega, A., 2000. Line-based, reduced memory, wavelet image compression. IEEE Trans. Image Process. 9 (3), 378–389.

ISO/IEC, 2000. JPEG2000 image coding system, JPEG2000 final committee draft v1.0, ISO/IEC JTC1/SC29/WG1 N1646.

Kovacevic, J., Sweldens, W., 2000. Wavelet families of increasing order in arbitrary dimensions. IEEE Trans. Image Process. 9 (3), 480–496.

Le Gall, D., Tabatabai, A., 1988. Subband coding of digital images using symmetric short kernel filters and arithmetic coding techniques. In: Proc. IEEE Internat. Conf. on Acoustics, Speech and Signal Processing, pp. 761–765.

Marcellin, M.W., Gormish, M.J., Bilgin, A., Boliek, M.P., 2000. An overview of JPEG-2000. In: Proc. IEEE Data Compression Conference, pp. 523–541.

Pennebaker, W.B., Mitchell, J.L., 1993. JPEG: Still Image Data Compression Standard. Van Nostrand Reinhold, New York.

Requirements AHG, 1999. JPEG2000 requirements and profiles v5.0, ISO/IEC JTC1/SC29/WG1 N1271.

Taubman, D., 1998. Report on coding experiment CodEff22: EBCOT (embedded block coding with optimized truncation), ISO/IEC JTC1/SC29/WG1 N1020R.

Taubman, D., 2000. High performance scalable image compression with EBCOT. IEEE Trans. Image Process. 9 (7), 1158–1170.