A STUDY ON NON-OCTAVE RESOLUTION CONVERSION BASED ON JPEG2000 EXTENSIONS

Takayuki Nakachi, Tomoko Sawabe, Junji Suzuki and Tetsuro Fujii

NTT Network Innovation Laboratories

ABSTRACT

The features JPEG2000, an international standard for still image compression, include 1) high coding performance, 2) unified lossless/lossy compression, 3) resolution and SNR scalability. Resolution scalability is an especially promising feature given the popularity of Super High Definition (SHD) images like digital-cinema. Unfortunately, its current implementation of resolution scalability is restricted to powers of two. In this paper, we introduce a non-octave resolution conversion method that is compatible with JPEG2000 part2. By using the proposed algorithm, images of various resolutions can be decoded from a compressed JPEG2000 part2 code stream. Experimental results from digital cinema test sequences show the effectiveness of the proposed algorithm.

1. INTRODUCTION

An ISO/ITU-T standard for still image coding was introduced as JPEG2000 [2]. The algorithm provides various features such as 1) high coding performance, 2) unified lossless and lossy compression, 3) resolution and SNR scalability. Since Super High Definition (SHD) images are becoming more popular [1], one application is digital-cinema, resolution scalability is regarded as a very promising attribute. The discrete wavelet transform used in JPEG2000 is dyadic and makes it easy to realize resolution scalability. Two-dimensional transformation is done by applying the one-dimensional wavelet transform to the rows and columns of an image. The same transformation is applied to the lowest subband iteratively. This multiresolution structure facilitates interchange among different video source formats. The wavelet transform is naturally scalable.

The current version of JPEG2000, however, restricts the resolution scalability to a power of two. In other words, the conversion rate is limited to $1/2^n$ ($n = 1, 2, \cdots, D$, where D is the decomposition level of the wavelet transform). Some applications in the areas of digital-cinema, HDTV, Standard TV, PC require a variety of image resolutions and so octave resolution scalability is too restrictive. Resolution conversion based on orthogonal transforms or a filter bank [4]-[6] realizes a non-octave resolution scalability. In particular, various DCT based resolution conversion methods have been discussed [4], [5]. DCT-type conversion methods have the advantage of being compatible with the JPEG image coding standard. However, this type of conversion method does not support JPEG2000. Reference [7] introduced a non-octave resolution conversion method that is compatible with JPGE2000. After a $1/2^n$ scale resolution image is processed by a JPEG2000 decoder, it is passed through a decimation filter. This method demands that the transmitted JPEG2000 code stream be redundant. If we want a $3/4$ scale resolution image, the JPEG2000 code stream corresponding to the full resolution image must be transmitted. In terms of coding efficiency, this is not a desirable characteristic.

In this paper, we introduce a non-octave resolution conversion method that is compatible with JPEG2000 part2. This method can output $N/2^D$ (N=1,2,$\cdots$,2^D) scale resolution images. The proposed method only need transmit JPEG2000 part2 code stream which corresponds to the $N/2^D$ scale resolution image while the method of S. Tanabe et al. transmits redundant data. It achieves non-octave resolution conversion by 1) simple N scale enlargement and 2) a wavelet low-pass filter which is used in JPEG2000 part1 or part2. Therefore, there is no need to design new decimation filters. Section 2 introduces the non-octave resolution conversion method. Section 3 shows simulation results. Conclusions are given in section 4.

2. NON-OCTAVE RESOLUTION CONVERSION METHOD BASED ON JPEG2000 PART2

This section describes the proposed non-octave resolution conversion method.

2.1. System Configuration

Encoder and decoder configurations are shown in Fig. 1; both are compatible with JPEG2000 part2. In the encoder, the input image is decomposed into uniform subbands, and coefficients of each subband are quantized. The quantized coefficients are then coded using the Embedded Block Coding with Optimized Truncation (EBCOT) algorithm [2]. The EBCOT algorithm realizes various levels of scalability. There

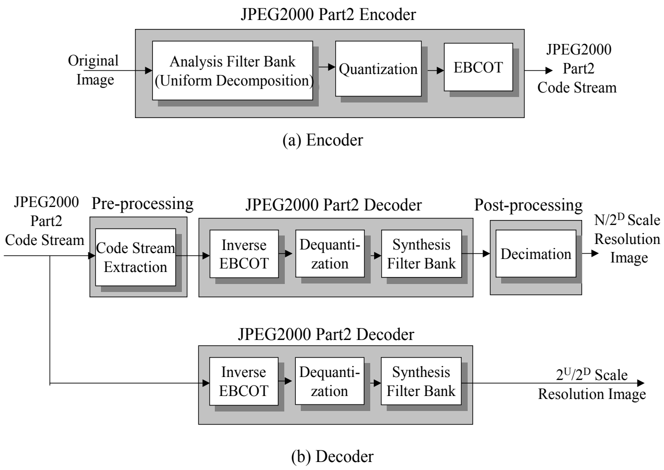

Fig. 1 System configuration.

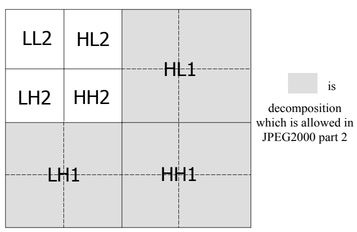

Fig. 2. Uniform subband decomposition of an image (Decomposition level $D = 2$).

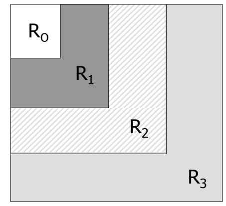

Fig. 3. Resolution levels $R_i (i = 0, 1, 2, 3)$ (Decomposition level $D = 2$).

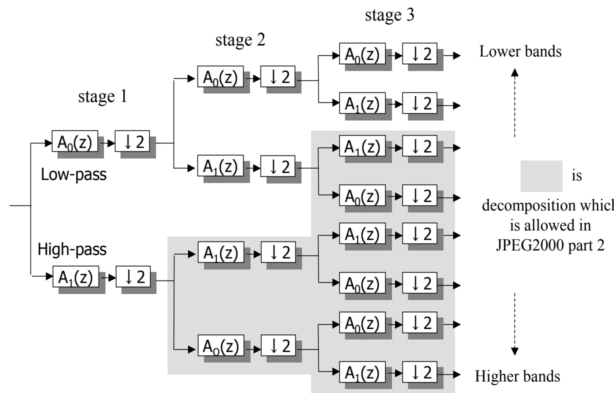

Fig. 4. Uniform subband decomposition by full tree-structured filter bank (Decomposition level $D = 3$).

are four basic scalability dimensions in a JPEG-2000 code stream: resolution (R), quality (L), spatial location (P), and component (C). Different scalability levels are achieved by ordering packets within the code stream. In order to achieve resolution scalability, we exploit the R-L-C-P order structure.

For subband decomposition, we use a full tree-structured filter bank, which yields uniform spectrum division. Arbitrary decomposition trees are allowed in JPEG2000 part2 while JPEG2000 part1 is restricted to Mallet decomposition. Figure 2 shows uniform subband decomposition of an image with two decomposition levels. In this case, the resolution level is defined as shown in Fig. 3. We define the resolution level from the lowest subband to the highest subband as $R_0, R_1, \cdots, R_{2^D-1}$.

In the decoder, a pre-processor extracts packets that correspond to the $R_0, R_1, \cdots, R_{N-1}$ resolution levels. The extracted code stream is then fed to a JPEG2000 part2 decoder. The JPEG2000 part2 decoder outputs the $2^U/2^D$ scale resolution image (where, U is a minimum integer value that satisfies the condition $N < 2^U$). In post-processing, the $2^U/2^D$ scale resolution image is converted into a $N/2^D$ scale resolution image.

When a JPEG2000 part2 code stream is directly decoded by a JPEG2000 part2 decoder, only octave scale resolution images can be output.

2.2. Analysis/Synthesis Filter Bank

We realize subband decomposition by using a full tree-structured filter bank; this yields uniform spectrum division. The wavelet filter, such as a reversible 5/3 or irreversible 9/7 filter, is used. For example, Fig. 4 shows uniform subband decomposition by the full tree-structured filter bank with three decomposition levels.

Upper side of Fig. 5 shows the relationship between the analysis filter bank and a synthesis filter bank. In the synthesis filter bank, the signal is reconstructed by summing the lower N subband signals. The U stages (where, U is the minimum integer value that satisfies the condition $N < 2^U$) are inevitable if we are to synthesis N subband signals. The result is an image with $2^U/2^D$ resolution. In the case of $N/2^D = 3/4$ scale resolution conversion, the amplitude spectrum of the synthesis filter bank output is as shown in Fig. 7. This shows that the frequency component of the highest subband is truncated.

2.3. Post-processing

In post-processing, the $2^U/2^D$ scale resolution image is converted to the $N/2^D$ resolution image. Although various

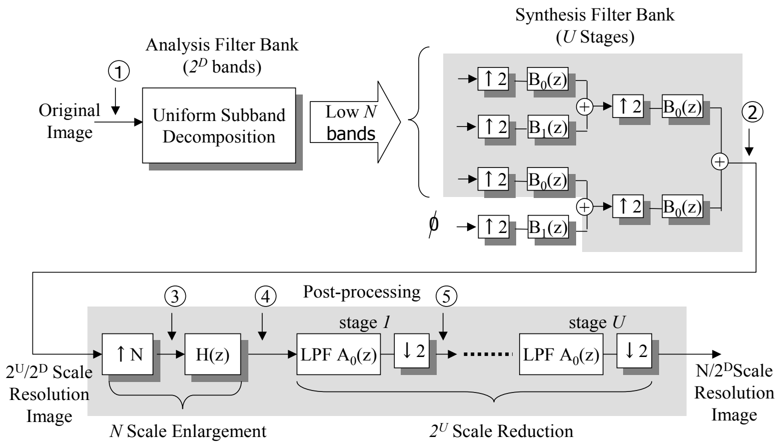

Fig. 5 Analysis/synthesis filter bank and post-processing.

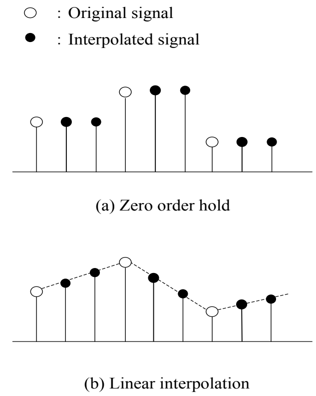

Fig. 6 N scale enlargement.

types of decimation filters can be used, we base resolution conversion on the following concepts; 1) simple N scale enlargement, and 2) decimation by a wavelet filter such as the reversible 5/3 or the irreversible 9/7 filter. As shown at the bottom of Fig. 5, $N/2^U$ scale resolution conversion is implemented by 1) N scale enlargement, and 2) $1/2^U$ scale reduction.

[N scale enlargement]

We implement N scale enlargement by 1) upsampling with factor N, 2) zero-order hold, or 3) linear interpolation. These are simple interpolation methods. Procedures of zero-order hold and linear interpolation are shown in Fig. 6. Either can be considered as a type of a low-pass filter $H(z)$ after upsampling with factor N. For example, the low-pass filter $H(z)$ of zero-order hold can be expressed by

$$H(z) = 1 + z^{-1} + \cdots + z^{-(N-1)}.$$

Figure 7 shows amplitude spectra of the upsampling, the zero-order hold, or the linear interpolation. Imaging components in the high frequency are caused by the upsampling. The low pass filter $H(z)$ suppress the imaging components.

[$1/2^U$ scale reduction]

$1/2^U$ scale reduction is achieved by using only the wavelet filter. Here, we use the reversible 5/3 or the irreversible 9/7 filter. Therefore, there is no need to design new decimation filters unlike ref. [7] which demands a different decimation filter for each scale.

Figure 7 shows an example of $1/4$ scale conversion. At first, the input signal is passed through the low-pass analysis filter. As a result, the high frequency image component is suppressed. The filtered output is then downsampled by factor 2, yielding low-pass subband signals. The resolution is reduced to half size. Repeating low-pass analysis filtering and downsampling U times yields $1/2^U$ scale reduction. Although aliasing is created by the downsampling process, the aliasing components are as small as possible. This is because the imaging components are suppressed by zero-order hold or linear interpolation in the process of N scale enlargement and their passage through the low-pass analysis filter in the $1/2$ scale reduction process.

3. SIMULATION RESULTS

Simulations were carried out to evaluate the effectiveness of the proposed method. We processed the original digital cinema test sequences called “Dance”, “Party”, ”Lake”, and “Aristocrat”: each with 1920×1024 [pixels], 8 [bits/pixel]. Reference images were created by DCT-IDCT conversion as shown in Fig. 8. At first, $K \times L$ DCT calculation was carried out for a $K \times L$ resolution image. Next, $K(N/2^D) \times L(N/2^D)$ IDCT is done for the lower $K(N/2^D) \times L(N/2^D)$ DCT components.

Table 1 shows the PSNR of $3/4$ resolution images. The analysis filter bank used the 5/3 reversible filter, and the decomposition level D was set to 2, N scale enlargement was realized by 1) upsampling with factor 3, 2) zero-order hold, and 3) linear interpolation.

$1/2^D$ reduction was implemented as 1) 9/7 irreversible filtering, and 2) 5/3 reversible wavelet filtering. From Table 1, linear interpolation gives the best performance among the three methods. It is considered that the aliasing of DC components is effectively suppressed by the linear interpolation. Compared to the reversible 5/3 filter, the irreversible 9/7 filter gives better performance.

Table 2 shows the PSNR values of the $5/8$ resolution images. The analysis filter bank used the 5/3 reversible filter, and the decomposition level D was set to 3. Table

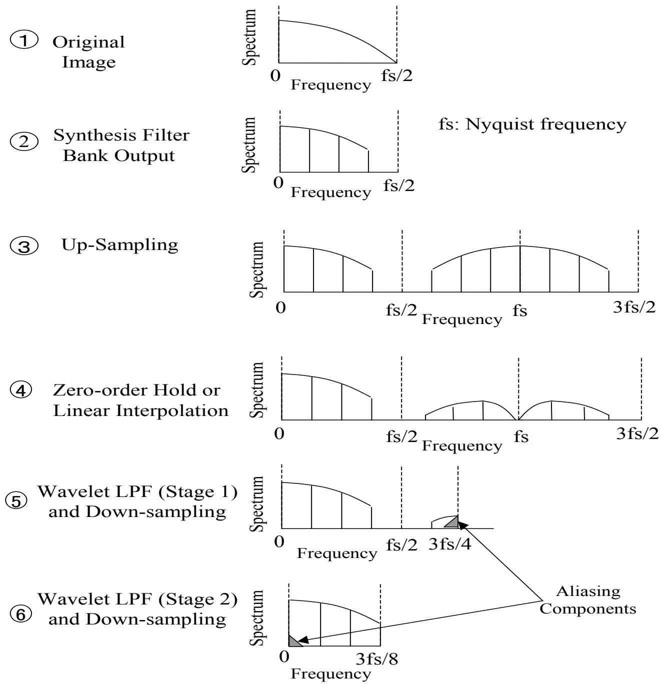

Fig. 7 Spectral interpretation of $3/4$ scale resolution conversion.

2 shows that zero order interpolation gave the best performance. This is a different result from that seen in $3/4$ scale conversion. Since the imaging and aliasing components were small in this case, zero order hold is very efficient.

4. CONCLUSIONS

This paper introduced a non-octave resolution conversion method compatible with JPEG2000 part2. The proposed algorithm allows $N/2^D(N = 1, 2, \cdots, 2^D)$ resolution images to be decoded from a compressed JPEG2000 part2 code stream. Experiments on digital cinema test sequences showed the effectiveness of the proposed algorithm. The features of the proposed resolution conversion method are summarized below:

1. Non-octave resolution scalability
2. Compatibility with JPEG2000 part 2
3. Implementation by simple enlargement and a wavelet filter

Evaluation of the coding performance of the proposed resolution conversion method coupled with EBCOT is a future study item.

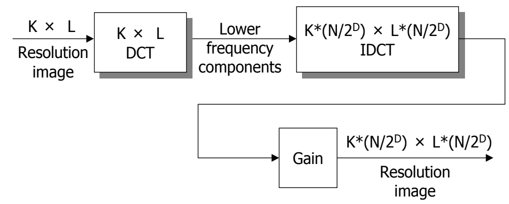

Fig. 8 Creation of reference image.

Table 1 PSNR[dB] of 1440 × 810 resolution images

(a) 9/7 irreversible filter

	Dance	Party	Lake	Aristo
Upsampling	30.36	24.62	27.02	24.66
Zero	38.40	33.62	32.67	28.85
Linear	40.94	37.39	37.26	34.34

(b) 5/3 reversible filter

	Dance	Party	Lake	Aristo
Upsampling	22.09	17.45	18.92	17.59
Zero	37.31	32.70	31.75	27.94
Linear	37.71	32.44	34.88	32.33

Table 2 PSNR[dB] of 1200 × 675 resolution images

(a) 9/7 irreversible filter

	Dance	Party	Lake	Aristo
Upsampling	34.87	28.92	30.96	27.88
Zero	36.83	31.85	30.89	26.91
Linear	33.04	27.40	29.42	26.65

(b) 5/3 reversible filter

	Dance	Party	Lake	Arist
Upsampling	22.69	17.77	19.29	17.57
Zero	36.05	30.78	29.77	25.77
Linear	32.95	27.18	29.63	27.02

5. REFERENCES

[1] T. Yamaguchi, et.al. "SHD Digital Cinema Distribution Over a Long Distance Network of Internet2," VCIP2003, Proc. SPIE, Vol.5150, pp.1760-1769, 2003.

[2] ISO/IEC 15444-1, *JPEG2000 Part I:Core coding system,* ISO/IEC JTC1/SC29 WG1, Dec. 2000.

[3] ISO/IEC FCD15444-2, *JPEG 2000 Part II Final Committee Draft,* ISO/IEC JTC1/SC29 WG1, Dec. 2000.

[4] Y. Miyamoto, Y. Akaiwa, "Image size conversion on DCT compressed data," IEICE Trans., D-II, vol. J82-D-II, no.1, pp.53-60, Jan. 1999.

[5] K. Kojima, H. Kiya, "Generalizations of the image resolution conversions in DCT-domain using 8 points inversed DCT," IEEE APCCAS, pp. 567-570, Nov. 1998.

[6] Y. Akutsu, H. Kobayashi, and H. Kiya, "A evaluation measure of resolution conversions with orthogonal transform, filter bank or wavelet transform", Technical Report of IEICE, DSP93-26, pp.17-24, May 1993.

[7] S. Tanabe, H. Watanabe, and H. Tominaga, "A study on image resampling for motion JPEG2000", Technical Report of IPSJ, AVM40-4, pp.19-23, Mar. 2003.