A STUDY ON NON-OCTAVE SCALABLE CODING USING MOTION COMPENSATED INTER-FRAME WAVELET TRANSFORM

Takayuki Nakachi and Tetsuro Fujii

NTT Network Innovation Laboratories
1-1 Hikari-no-oka Yokosukaishi, Kanagawa, 239-0847 JAPAN
e-mail: nakachi.takayuki@lab.ntt.co.jp

ABSTRACT

JPEG2000, an international standard for still image compression, has three main features: 1) high coding performance, 2) unified lossless/lossy compression, and 3) resolution and SNR scalability. Resolution scalability is especially promising given the popularity of super high definition (SHD) images like digital-cinema. Unfortunately, the resolution scalability of its current implementation is restricted to powers of two. In this paper, we introduce non-octave scalable coding with motion compensated inter-frame wavelet transform. By using the proposed algorithm, images of rational scales can be decoded from a compressed code stream. Experiments on SHD 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 [1]. The algorithm provides various features but the key points are 1) high coding performance, 2) unified lossless and lossy compression, and 3) resolution and SNR scalability. Since Super High Definition (SHD) images are becoming more popular [2], resolution scalability is particularly attractive. The discrete wavelet transform used in JPEG2000 is dyadic which makes it easy to realize resolution scalability.

One problem with JPEG2000 is that its resolution scalability is restricted powers 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, and PCs require a variety of image resolutions and so octave resolution scalability is too restrictive. Resolution conversion based on orthogonal transforms or a filter bank [3]-[4] can realize non-octave resolution scalability.

DCT-type conversion methods have the advantage of being compatible with the JPEG image coding standard. However, this type of conversion method does not possess other necessary properties such as SNR scalability.

Recently, we introduced non-octave scalable coding (NSC) [5] that uses a filter bank with embedded block coding with optimized truncation (EBCOT) [1]. This method can output $N/2^D$ (N=1,2,$\cdots$,2^D) scale resolution images and also possess SNR scalability. In this paper, we propose a new type of non-octave scalable coding which improves upon the coding efficiency of our earlier NSC. The new method utilizes motion compensated inter-frame wavelet transform while NSC uses only intra frame subband transform. Section 2 introduces the new method called motion compensated non-octave scalable coding (MC-NSC). Section 3 shows simulation results. Our conclusions are given in section 4.

2. NON-OCTAVE SCALABLE CODING USING MC INTER-FRAME WAVELET TRANSFORM

This section describes non-octave scalable coding with motion compensated inter-frame wavelet transform as an extension of our earlier NSC.

2.1. System Configuration

Encoder and decoder configurations are shown in Figs. 1 and 2, respectively. In the encoder, the input image is decomposed into 2^D uniform subbands. Motion compensated inter-frame coding is applied only on the lowest LL subbands; the remaining higher subbands are coded in intra-frame mode. This is because the inter-frame correlation of higher subbands is relatively weak, especially in digital cinema sequences. The inter/intra frame coefficients are separately coded using a embedded entropy coding scheme. The embedded coding algorithm realizes various levels of scalability. There are four basic scalability dimensions: spatial resolution (SR), temporal resolution (TR), quality (L), and component (C). Different scalability levels are achieved by ordering packets within the code stream. Details of inter-frame coding are described in Sec. 2.3. The intra frame coding procedure is the same as that used in our earlier NSC.

In the decoder, a pre-processor extracts packets that correspond to the lower N subbands. This yields $N/2^D$ scale

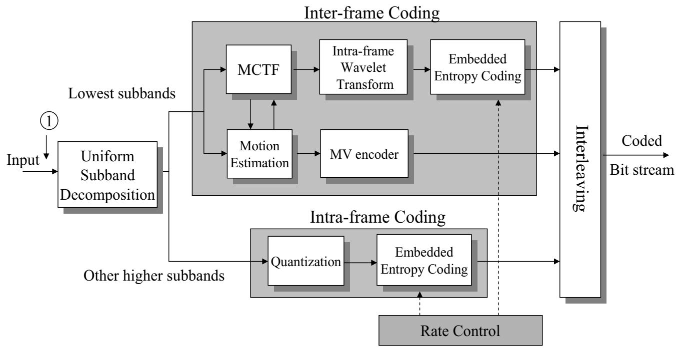

Fig. 1. Encoder configuration.

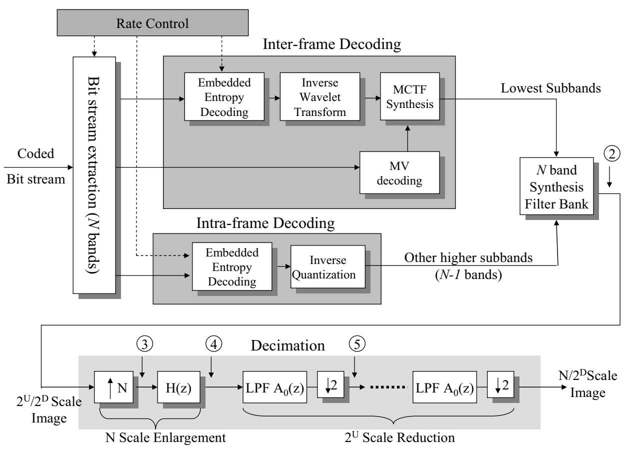

Fig. 2. Decoder configuration.

conversion. The extracted code corresponding to the LL subband is then fed to the inter-frame decoder, and the remaining $N - 1$ packets are fed to the intra-frame decoder. Finally, a synthesis filter bank reconstructs the signals by summing the output signals of the inter-frame and intra-frame decoders. The synthesis filterbank outputs $2^U/2^D$ scale resolution images (where, U is the minimum integer value that satisfies the condition $N < 2^U$).

Post-processing converts the $2^U/2^D$ scale resolution image into a $N/2^D$ scale resolution image.

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. 3 shows uniform subband decomposition by a full tree-structured filter bank with three decomposition levels.

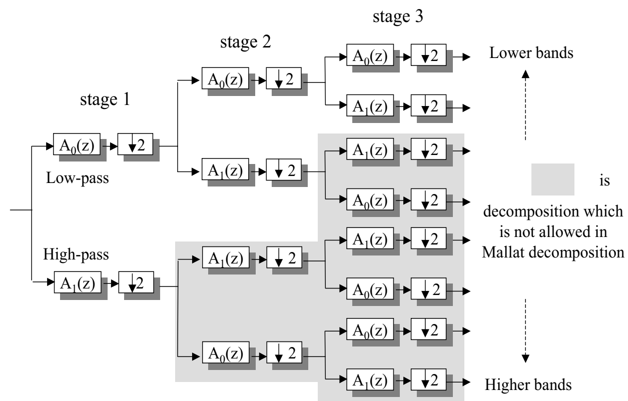

Fig. 3. Uniform subband decomposition by full tree-structured filter bank (Decomposition level $D = 3$).

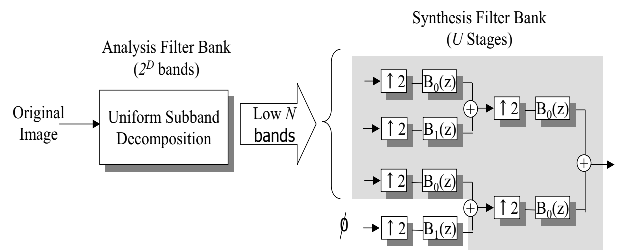

Fig. 4. Relationship between analysis/synthesis filter bank.

Figure 4 shows the relationship between the analysis filter bank and the 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. 5. This shows that the frequency component of the highest subband is truncated.

2.3. Motion Compensated Inter-frame Wavelet Coding

Motion compensated time filtering (MCTF) [6] is imposed on lowest subbands. MCTF plays an essential role in exploiting the temporal redundancy in image sequence. MCTF is performed along the motion trajectory. A pair of temporal low and high subbands is generated for each successive input signal by the wavelet transform. As shown in Fig. 6, the operation is recursively performed on the low temporal frequency subband. Hierarchical variable size block match-

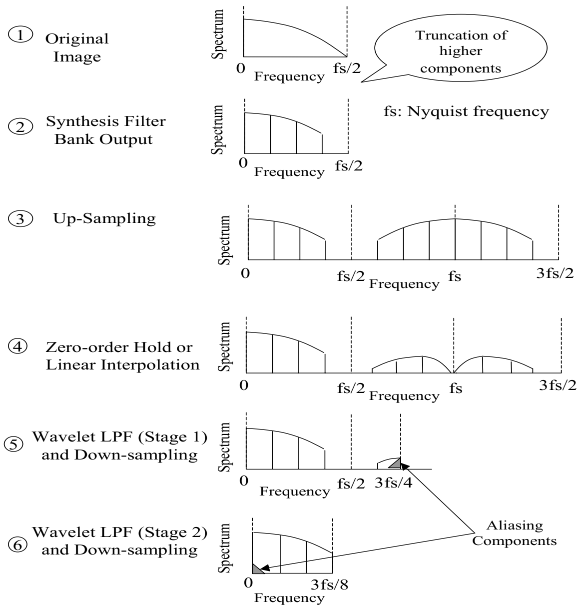

Fig. 5. Spectral interpretation of $3/4$ scale resolution conversion.

ing (HVSBM) is employed for motion estimation since it speeds up the search process and spatial interaction is imposed between neighboring motion vectors. The sizes of the square motion blocks range from 4×4 to 64×64.

Spatial wavelet transform follows MCTF to complete 3D subband decomposition. The numbers of spatial and temporal decomposition levels will vary with input image resolution and delay constraints. As a result, the inter-frame coding divides consecutive frames into groups of pictures (GOP). The 3D subband structure in a GOP is depicted in Fig. 7. Finally, each 3D subband is separately coded using the embedded entropy coding scheme. The coded data are packetized in terms of scalability; spatial resolution (SR), temporal resolution (TR), quality (L) and component (C). Different scalability levels are achieved by ordering packets within the code stream.

2.4. Decimation Process

Post-processing converts the $2^U/2^D$ scale resolution image to the $N/2^D$ resolution image. As shown at the bottom of Fig. 2, we implement $N/2^U$ scale resolution conversion by 1) N scale enlargement, and 2) $1/2^U$ scale reduction. The N scale enlargement is implemented by 1) up sampling with factor N, 2) zero-order hold, or 3) linear interpolation. These are simple interpolation methods and all can be considered as a type of a low-pass filter $H(z)$ after up sampling with factor N. Figure 5 shows the amplitude spectra yielded by up sampling, zero-order hold, and the linear interpolation.

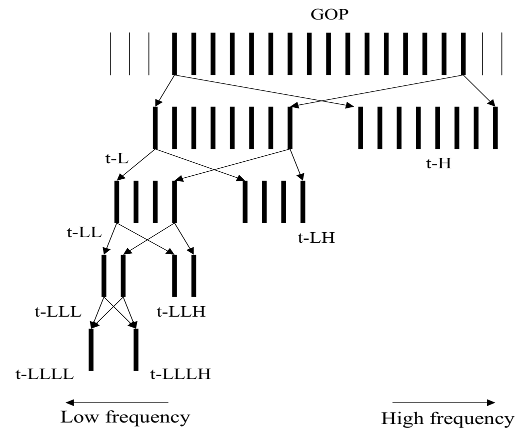

Fig. 6. Octave based temporal decomposition [6].

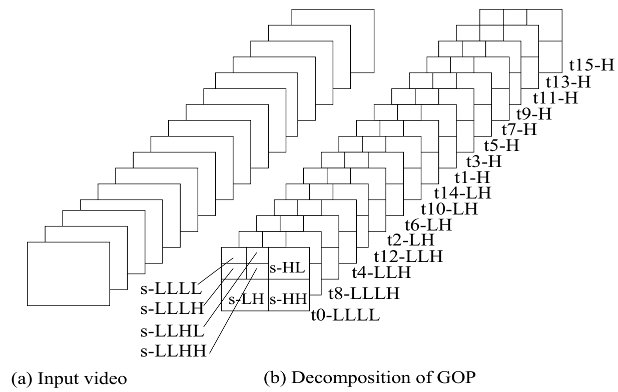

Fig. 7. 3D subband decomposition [6].

The imaging components in the high frequency region are caused by up sampling, and are suppressed by the low pass filter $H(z)$.

$1/2^U$ 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. Figure 5 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 down sampled by the factor of 2, yielding low-pass subband signals. The resolution is reduced to half size. Repeating the cycle of low-pass analysis filtering and down sampling U times yields $1/2^U$ scale reduction.

3. SIMULATION RESULTS

Simulations were carried out to evaluate the effectiveness of the proposed method. We processed a digital cinema test sequence consisting of 4096×1969 [pixels], 8 [bits/pixel], 24[fps]. Reference images were created by DFT-IDFT con-

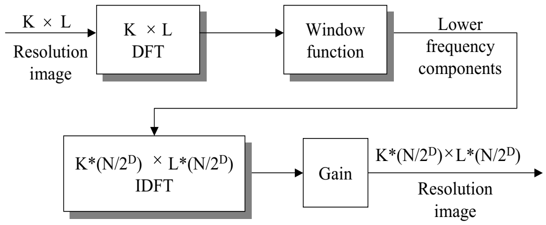

Fig. 8. Creation of reference image.

version as shown in Fig. 8. At first, $K \times L$ DFT calculation was carried out for a $K \times L$ resolution image. Next, by applying a window function to the lower $K(N/2^D) \times L(N/2^D)$ DFT signals, $K(N/2^D) \times L(N/2^D)$ signals can be extracted. At last, $K(N/2^D) \times L(N/2^D)$ IDFT was done for the extracted $K(N/2^D) \times L(N/2^D)$ DFT signals.

Here, we compare the proposed model to intra non-octave scalable coding NSC. In both the proposed method and NSC, 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 linear interpolation and $1/2^D$ reduction was implemented as 5/3 reversible wavelet filtering. In the proposed method, each GOP contains 8 frames: one t-LLL frame, one t-LLH, two t-LH frames and four t-H frames. Motion vector accuracy was a quarter pixel.

Figure 9 shows the rate-distortion characteristic for a full scale resolution sequence. The proposed method (MC-NSC) has better performance than NSC. Figure 10 shows the rate-distortion characteristic for a $3/4$ scale resolution (non-octave resolution) sequence. These results show that the proposed method is superior to NSC, especially at lower bit-rates.

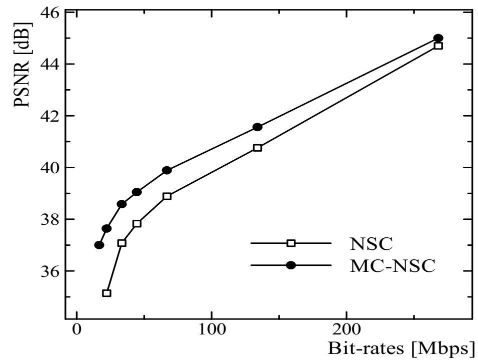

Fig. 9. Rate-distortion for a full resolution sequence.

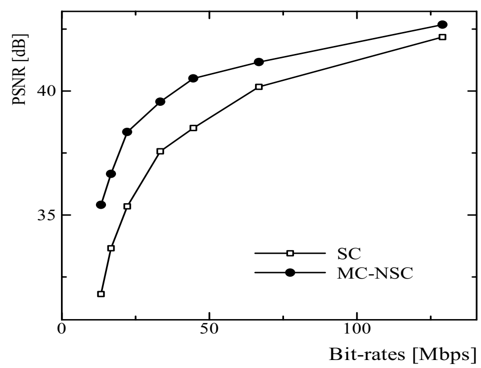

Fig. 10. Rate-distortion for a $3/4$ scale resolution sequence.

4. CONCLUSIONS

This paper introduced motion compensated non-octave scalable coding (MC-NSC). The proposed algorithm allows $N/2^D$ ($N = 1, 2, \cdots, 2^D$) resolution images to be decoded from the same code stream. Experiments on digital cinema test sequences showed that MC-NSC outperforms intra NSC. The features of the proposal are summarized below:

1. Non-octave resolution scalability
2. SNR/temporal scalability
3. High coding performance when compared with NSC

The following points are left as future studies:

1. Applying MCTF to higher subbands.
2. Consideration of MC-NSC coded data transmission over IP networks.

5. REFERENCES

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

[2] T. Yamaguchi, et al. ”SHD digital cinema distribution over a long distance network of internet2,” VCIP2003, Proc. SPIE, Vol.5150, pp.1760-1769, 2003.

[3] Y. Miyamoto, et al., ”Image size conversion on DCT compressed data,” IEICE Trans., D-II, vol. J82-D-II, no.1, pp.53-60, Jan. 1999.

[4] Y. Akutsu, et al., ”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.

[5] T. Nakachi, et al., “A study on non-octave resolution conversion Based on JPEG2000 Extensions,” IEEE ICASSP2004, III-69, May 2004.

[6] S. T. Hsiang, et al., “Invertible temporal subband/wavelet filter banks with half-pixel-accurate motion compensate,” Motion-compensated 3-D subband coding of video” IEEE trans. Image Processing, No. 8, Aug. 2004.