Interleaved Multiple Frame Coding using JPEG2000

Ayuko Takagi and Hitoshi Kiya
Dept.of Electrical Engineering, Tokyo Metropolitan Univ.
1-1 Minami-ohsawa Hachioji-shi,Tokyo 192-0397,JAPAN
Tel: +81-426-77-2745, Fax: +81-426-77-2756
E-mail: ayuko@isys.eei.metro-u.ac.jp, kiya@eei.metro-u.ac.jp

Abstract: This paper describes an effective technique for coding video sequences based on JPEG2000 codec. In the proposed method, multiple frames are combined into one large picture by interleaving each pixel data. A large picture enables images to be coded more efficiently and image quality is improved. A video sequence is efficiently coded by adapting the time correlation of the video sequences to spatial correlation. We demonstrated the effectiveness of this method by encoding video sequences using JPEG2000.

1. Introduction

In most multimedia applications, video data is compressed by either of the following two methods: the inter-frame coding, and the intra-frame coding. It is known that the neighboring frames have a high correlation in video sequences. Inter-frame coding, which is used in the MPEG codec, uses motion estimation to eliminate the temporal redundancy of the video frames, thus attaining a high rate of compression. However, motion estimation is the most time-consuming task in MPEG encoder[1]. In contrast, intra-frame coding has more ease on its coding process in comparison to inter-frame coding, and each video frame is encoded independently. Since the temporal correlation of video sequences is not used, intra-frame coding has lower coding efficiency compared to the inter-frame coding. In general, when the video data is encoded on the same bitrate, the image quality of inter-frame coding becomes higher than that of intra-frame coding. For lossless video coding, the bitrate is reduced by using inter-frame coding, and algorithms using motion estimation have been studied[2], [3].

Codecs using intra-frame coding include JPEG, JPEG2000 [4], and MotionJPEG2000[5]. JPEG2000 and Motion-JPEG2000, has been standardized recently. JPEG2000 enables more efficient coding than JPEG, and also, lossy and lossless coding can be done by the same encoder. However the coding efficiency decreases as the picture size becomes small. By using a large picture, the coding efficiency increases for JPEG2000[6].

In this paper, we propose an effective technique for efficiently coding video sequences using JPEG2000. In the proposed method, we generate a large picture by interleaving multiple frames. The coding efficiency increases by using the generated large picture. This proposed method are done in advance of encoding, and does not modify the codec.

2. JPEG2000

2.1 JPEG2000 encoder

JPEG2000 is based on wavelet transform, and it enables more efficient coding than JPEG, which is based on DCT (discrete cosine transform). Furthermore, JPEG2000 treats both lossy and lossless coding with the same encoder. Here, we describe an outline of the JPEG2000 encoder.

Firstly, input images are decomposed into several sub-bands by applying the DWT (Discrete Wavelet Transform). The lifting-based DWT is adopted in JPEG2000. There are two forms of lifting-based DWT. One is the lossless DWT which is called the integer wavelet transform(IWT), the other is the lossy DWT which is referred to as the floating-point wavelet transform (FWT). Note that 'lossless transform' as described here allows perfect reconstruction to integer accuracy even though the transformed coefficients are also to integer accuracy. One of the advantages of the lifting-based DWT is that this technique allows the unification of lossless and lossy coding. For lossy coding, the wavelet coefficients of each sub-band are quantized. Then, each sub-band coefficients are divided into non-overlapping rectangles, called code-blocks, which form the input to the entropy coder.

2.2 Relationship between image size and compression rate

In JPEG2000, when the picture size is reduced, compression efficiency decreases. This occurs due to decrease in frequency resolution.

In wavelet transform, the pixel data is passed through a lowpass and highpass filter, then down sampled by a factor of two, constituting one level of transform. Multiple levels of the wavelet transform are made by repeating the filtering and decimation process on the lowest sub-band components. Let L denote the total number of levels of decomposition. When the image size is $M \times N$, the the size of lowest sub-band components (LL) will be $\frac{M}{2^L} \times \frac{N}{2^L}$. Figure 1 shows an example of three level decomposition. In order to maintain LL components at a certain effective size, there are limitations on the decomposition level L according to the image size. That is, frequency resolution is low for small pictures. Also, in JPEG2000, the efficiency of arithmetic coding of the LL component is statistically good compared with that of the components of other sub-bands. With the same number of decomposition level, the size of LL components is proportional to the size of the input image. Therefore, with a small picture, the size of LL component is also small and coding efficiency decreases relatively.

To demonstrate the effects of picture size on coding efficiency, we show the results of a simulation using JPEG2000. The picture used was one frame of video sequence "mobile & calendar"; the picture size was 704×576 pixels. The simulation conditions were as follows:

1. The whole picture was encoded at once.

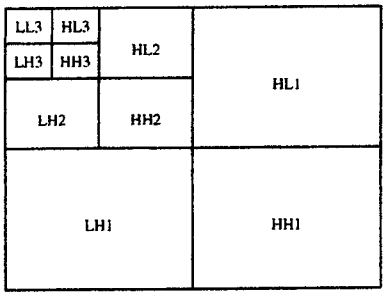


Figure 1. Example of three level 2-D wavelet transform.

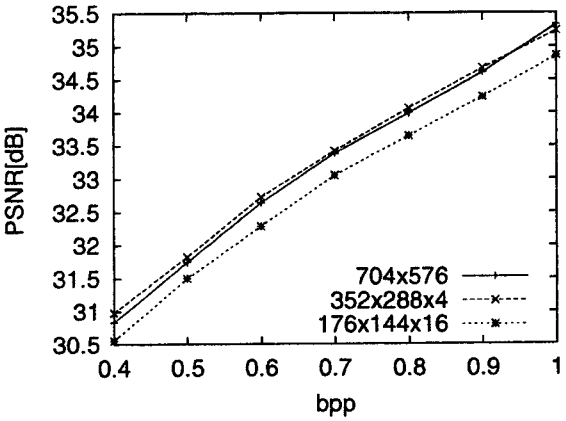


Figure 2. Influence of picture size on image quality in JPEG2000.

2. The picture was divided into four sections (352×288 pixels) and each section was encoded individually.
3. The picture was divided into 16 sections (176×144 pixels) and each section was encoded individually.

The encoding conditions for JPEG2000 were as follows: the wavelet transform was performed in three stages; the code block size was 64×64. The encoding bit rate varied from 0.4 to 1.0 [bpp]; Figure 2 shows the average value of PSNR for each bit rate. At the same bit rate, the highest image quality was equivalent to the maximum coding efficiency. The results showed that there was a slight difference between the image quality obtained with a picture size of 704 × 576 and that with a picture size of 352×288. However, image quality was lower with a picture size of 176×144. The level of image degradation was about 0.5 [dB]. That is, coding efficiency for data for small pictures decreased.

3. Proposed method for more efficient coding

3.1 Method for efficient coding

Here, we describe a method for efficiently coding small video sequences with using JPEG2000. Given the small picture size, it is not efficient to use JPEG2000 to encode each frame of a video as it is. However, the problem of decreased coding efficiency, outlined in Section 2, can be reduced by using a large picture. Therefore, we generate a large picture from several frames and use this as an input. Figure 3 shows the coding procedures for both the conventional method and the proposed method. In the conventional method, Figure 3 (a), each frame is encoded individually by the JPEG2000 encoder. Figure 3 (b) shows the proposed method which generates the large picture using four frames. Here, we generate a large picture by simply placing four frames side by side. The generated picture is then used as an input for the JPEG2000 encoder. Coding is more efficient since a large picture is used as an input. After decoding, a reverse procedure is performed and decoded pictures of the original size are obtained by dividing the large decoded picture. The resulting pictures have higher image quality than when each frame is coded individually.

3.2 Proposed method for generating large picture

Here, we show the method to generate a large picture from multiple frames. The large picture is generated by interleaving the pixel data from each frames. In video sequences, temporal correlation between adjacent frames are high. By combining these frames, the temporal correlation is transformed into the spatial correlation. There are flexibilities on the number of frames to use and how to arrange the pixels for the generation of large picture. Figure 4 shows an example of generating the large picture using four frames. The interleaving is performed per pixel of each frame. The image quality after the coding is effected by the correlation between neighboring pixels of the generated picture, and it differs by the number of frames to use and the arrangement of the pixel. Moreover, the temporal correlation is transformed into the spatial correlation, so the proposed method is less effective on video sequences with complex motion.

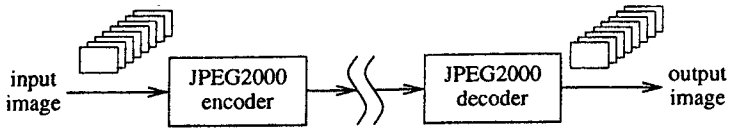
When using the proposed method, the arrangement procedure is needed for several frames in advance. This causes a delay in the encoding scheme. Also, when the number of frames becomes larger for generating a large picture, a larger amount of memory is needed for both encoding and decoding. In the decoding procedure, the memory is needed for buffering the decoded large picture, and each frame is rearranged and is used as an output. Although there are disadvantages, the image quality improves on using the proposed method.

4. Simulation

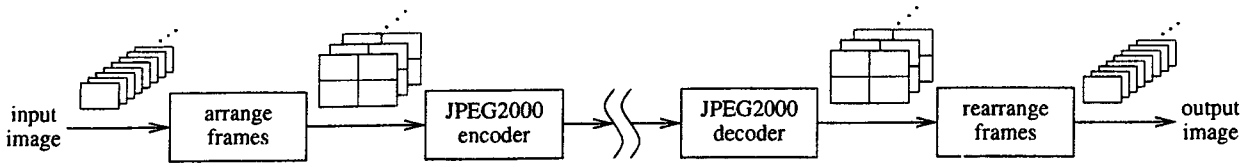
We show the results of a simulation to test the effectiveness of the proposed method. The test video sequences are "carphone","news","silent " (size: 176 × 144), "tempete", "paris" (size: 352 × 288), and "mobile & calendar" (size: 720 × 576). They were all grayscale sequences. The number of frames used was 32. For JPEG2000, the number of stages used for wavelet transform was three, and the size of code block was 64 × 64. When using the proposed method, we used four and sixteen frames to generate the large picture. The header size of JPEG2000 bitstream was 84 bytes for all results.

4.1 Lossless coding

For comparison, the video sequences were encoded using both proposed and conventional with JPEG-LS[7], JPEG2000[8] codecs. Table 1 and 2 gives the results



(a) Conventional method



(b) Proposed method for arranging four frames

Figure 3. Coding procedure.

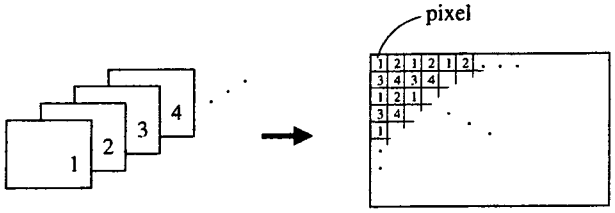


Figure 4. The example of generating large picture from four frames. The interleaving is performed per pixel of each frame.

Table 1. The bitrate comparison using JPEG-LS codec between the conventional and the proposed method showing the average bitrate of 32 frames [bpp].

codec	JPEG-LS		
method	conv.	proposed	
frames	1	4	16
carphone	3.933712	3.894860	3.904803
silent	4.672697	3.878512	3.458406
news	4.254886	3.237492	2.774036
paris	4.807087	4.140692	3.906511
tempete	5.154060	5.156425	5.331097
mobile	4.658587	5.309905	5.527094
average	4.580172	4.269648	4.150325

of the simulation for lossless coding using JPEG-LS and JPEG2000. For video sequences "carphone", "silent", "news", and "paris", the proposed method achieved high rate of compression. However, the proposed method was less effective for "tempete" and "mobile & calendar". In these two video sequences, there are panning movements in camera motion, which causes complex movements compared to others. Consequently, the temporal correlation of adjacent frames are low for these two video sequences, and the proposed method is less effective. The proposed method is more effective on small pictures for JPEG2000.

Table 2. The bitrate comparison using lossless JPEG2000 between the conventional and the proposed method showing the average bitrate of 32 frames [bpp].

codec	JPEG2000		
method	conv.	proposed	
frames	1	4	16
carphone	4.747619	4.468960	4.503939
silent	5.173184	4.622534	4.262666
news	5.136791	4.428543	3.837877
paris	5.477549	4.967056	4.734180
tempete	5.646563	5.341715	5.579545
mobile	4.862243	4.861489	5.377709
average	5.173992	4.781716	4.715986

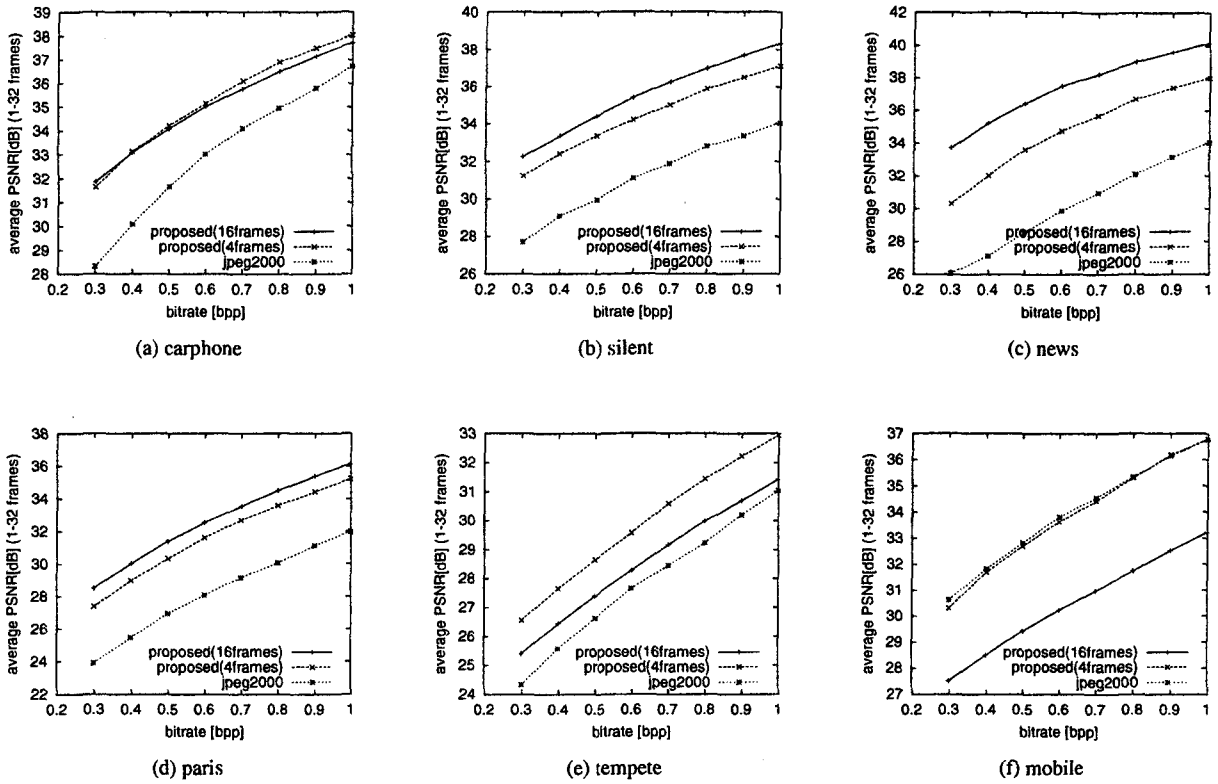
4.2 Lossy coding

Figure 5 shows the result of simulation using JPEG2000. The range of the coding bitrate is 0.3 – 1.0 [bpp]. The same tendency is observed between lossy and lossless compression. The proposed method is less effective on video sequences "mobile & calendar" (Figure 5(f)). For video sequence "tempete" (Figure 5(e)), the image quality is degraded for the proposed method using 16frames. The number of frames to use vary according to the characteristic of the video sequence. When there are complex motion in video sequence, the number of frames to use should be small.

The above results for lossless and lossy coding show that the proposed method is effective on video sequences with less movements. For video sequences with complex motion, motion estimation is needed to reduce temporal correlation effectively.

5. Conclusion

We have described a technique for coding video sequences efficiently using JPEG2000. Multiple frames are combined to form one large picture, and coding efficiency is improved by transforming the temporal correlation to spa-



(a) carphone (b) silent (c) news

(d) paris (e) tempete (f) mobile

Figure 5. The PSNR comparison between the conventional and the proposed method showing the average PSNR of 32 frames for each bitrate [dB].

tial correlation. The proposed method is effective on video sequences with less movements, and it has been shown by coding video sequences using JPEG-LS and JPEG2000.

References

[1] B. Furht, J. Greenberg, and R. Westwater,"Motion Estimation Algorithms for Video Compression", Kluwer Academic Publishers, 1997

[2] N.Ono and Y.Yashima, "An Efficiency Evaluation of Lossless Video Coding Using Motion Compensation," Technical Report IE95-18, IEICE, pp.21–27, May 1995.(in Japanese)

[3] G.C.K.Abhayaratne and D.M.Monro, "Embedded to lossless codimg of motion compensated prediction residuals in lossless video coding," Visual Communications and Image Processing 2001, Proc. SPIE vol.4310, pp.175–185, San Jose, CA, 21-26 January 2001.

[4] ISO/IEC IS 15444-1: "Information technology – JPEG2000 image coding system – Part 1: Core coding system", 2000

[5] ISO/IEC JTC1/SC29/WG1 N1987: "Motion JPEG2000 Committee Draft 1.0",ISO/IEC JTC1/SC29/WG1, Dec. 2000

[6] A. Takagi, M. Kurosaki, and H. Kiya, "The Effective method for Applying JPEG2000 to QCIF Video Coding", Technical Report IE2001-160, IEICE, Jan. 2002

[7] ISO/IEC JTC1/SC29/WG1 FCD14495: "Lossless and near-lossless coding of continuous tone still images(JPEG-LS)", ISO/IEC JTC1/SC29/WG1, Jul.1997.

[8] ISO/IEC JTC 1/SC29/WG1 WG1N1894: "JPEG2000 Verification Model 8.6," 2000