LETTER Special Section on Papers Selected from ITC-CSCC 2002

Lossless Video Coding Using Interleaved Multiple Frames

Ayuko TAKAGI[†a)] ***and*** **Hitoshi KIYA**[††]**, *Regular Members***

SUMMARY An effective method for lossless video compression using intra-frame coding algorithms is described. Intra-frame coding is mainly used for still image compression and does not exploit a temporal correlation in video sequences. With the proposed method, multiple frames are combined into one large picture by interleaving all pixel data. In a large generated picture, the temporal correlation of the video sequence is transformed into a spatial correlation. A large picture enables images to be coded more efficiently and decreases the bitrate for lossless coding using intra-frame coding. We demonstrated the effectiveness of this method by encoding video sequences using JPEG 2000 and JPEG-LS.

key words: *JPEG 2000, Motion JPEG 2000, lossless video coding, interleaved multiple frames*

1. Introduction

The necessity for lossless video coding has been increasing [1]–[5]. Lossless compression is required for archiving digital video content. Many technologies for lossless compression have been studied for still image coding [7]–[11], and some algorithms and encoders have been used. For lossy compression, one of two methods are used for video coding: inter- or intra-frame coding. In general, however, intra-frame coding has a lower coding efficiency compared with inter-frame coding.

Researchers have described lossless compression for intra-frame coding [7]–[11]. These techniques are for still image coding and exploit the spatial correlation of still images. To encode video sequences using techniques for still images, each video frame is coded independently. Because the temporal redundancy of video sequences is not used, the coding efficiency is decreased. JPEG-LS [7] and JPEG 2000 [8] were introduced for the still image coding standard and are in wide use. JPEG 2000 offers both lossy and lossless compression at the same encoder, and Motion JPEG 2000 [12] is being discussed for video coding.

Neighboring frames are known to have a high correlation in video sequences. For lossless video coding, the bitrate is reduced by using inter-frame coding. Inter-frame coding algorithms using motion compensation have already been studied [1]–[6]. In the literature, motion compensation was applied for efficient video coding. In one case, MPEG-2 was performed as a basic operation, and residual image coding offered an enhancement layer [2]. In another case, a wavelet transform was used as a residual image for embedded coding [4]. In Ref. [5], resolution scalability was achieved using a wavelet transform, and motion compensation was performed in the wavelet transformed domain. Motion compensation was used to predict the pixel information from the previous frame in reference [6]. Motion compensation contributes to the coding efficiency, but the process is time consuming and makes it difficult for editing video sequences [13].

In this paper, we propose an effective technique for lossless video coding that does not use inter-frame coding, but the coding efficiency increases compared with intra-frame coding. Our goal is to exploit the temporal redundancy for video coding without using motion compensation. In the proposed method, we do not use motion compensation, but we combine multiple frames into one by interleaving pixel data from each frame. Consequently, the temporal correlation of the video sequences is transformed into a spatial correlation, and in most cases, the coding efficiency is increased over the efficiency of coding each of the frames independently. This process is applied before encoding and does not modify the codec.

2. Proposed Method Using Interleaved Multiple Frames

Video sequences have a high temporal correlation, and inter-frame coding is used to exploit the temporal redundancy to achieve a high compression rate. Inter-frame coding requires a time consuming process, such as motion estimation, to remove the temporal redundancy effectively. In contrast, the encoding process is simple for intra-frame coding, which only uses the spatial correlation. In general, intra-frame coding has a smoother coding process than inter-frame coding because the coding concludes within a single frame. However, the coding efficiency is inferior because the temporal redundancy is not exploited.

Our goal is to increase the coding efficiency for video compression using an intra-frame coding algorithm. Using a temporal correlation is important to encode video sequences more efficiently.

2.1 Transform Temporal to Spatial Correlation

In Ref. [14], a method using a large picture was described. Figure 1 shows the coding procedure for the conventional method using JPEG 2000. The generated picture is then

Manuscript received December 24, 2002.
Final manuscript received February 6, 2003.
[†]The author is with the Department of Informatics and Media Technology, Shohoku College, Atsugi-shi, 243-8501 Japan.
[††]The author is with the Department of Electrical Engineering, Tokyo Metropolitan University, Hachioji-shi, 192-0397 Japan.
a) E-mail: a-takagi@shohoku.ac.jp

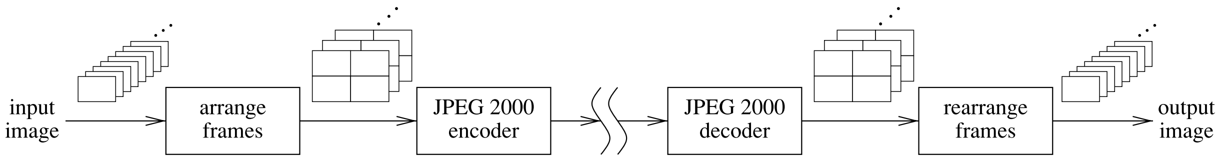

Fig. 1 Conventional method that generates a large picture as an input for the JPEG 2000 encoder.

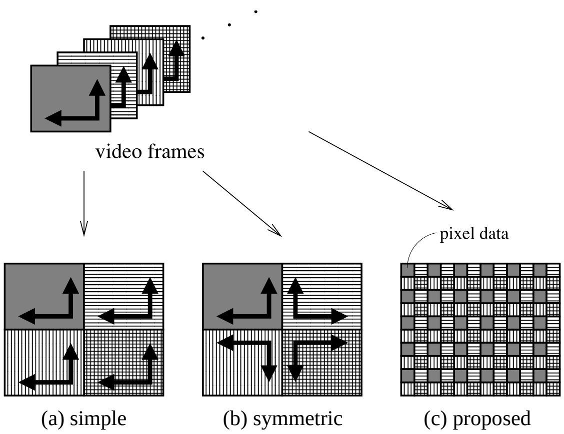

Fig. 2 Methods for generating a large picture from four frames.

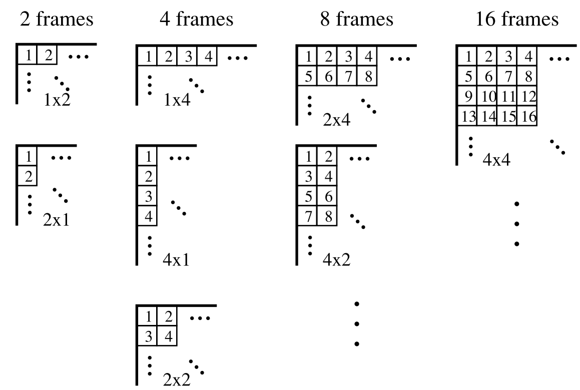

Fig. 3 Example of generating a large picture from 2, 4, 8, and 16 frames. Interleaving is performed for each pixel of each frame.

used as an input for the JPEG 2000 encoder. After decoding, a reverse procedure is performed and decoded pictures of the original size are obtained by dividing up the large decoded picture. By combining multiple frames into one large picture, the temporal correlation between the frames is transformed into a spatial correlation within the large generated picture. There are several ways to generate a large picture. Figure 2 shows an example of generating one from four frames. The simplest way is to place frames side by side as shown in Fig. 2(a). In Ref. [14], they generated a large picture by placing frames symmetrically as in Fig. 2(b) to smoothen the boundary between adjacent frames.

2.2 Interleaved Multiple Frames

We propose a method to exploit the temporal correlation of video sequences in intra-frame coding. The proposed method combines adjacent frames into one large picture and uses it as an input for the intra-frame encoder. The method for generating a large picture mentioned above is not sufficient to improve the correlation. In video sequences, the temporal correlation between adjacent frames is high. Moreover, the pixel data at the same position has a high correlation among different frames. Placing the pixel data at the same position close to each other is better. Here, we show how to generate a large picture from multiple frames by interleaving the pixel data from each frame. An example of generating a large picture from four frames is shown in Fig. 2(c). Interleaving is performed for each pixel of each frame. This way, the neighboring pixels are from the same position, and the spatial correlation improves. When the temporal correlation is sufficiently high, the spatial correlation of the generated frame becomes higher than that of the original frames. In this case, the coding efficiency increases. In addition, the size of the generated frame increases, therefore, the coding efficiency is improved.

2.3 Flexibilities

The pixel arrangement and the number of frames are flexible. First, we consider the arrangement. Figure 3 shows an example of generating a large picture using 2, 4, 8, and 16 frames. For example, when using four frames, the frames are arranged in three different ways, with the pixel data from each frame placed horizontally (1×4), vertically (4×1), or in a square (2×2). Each pixel of each frame is interleaved individually. The coding efficiency is affected by the correlation between the neighboring pixels of the generated picture, and it differs depending on the arrangement. Based on the spatial correlation of video frames, we should change the way pixel data is arranged in order to improve the correlation coefficient. Next, the number of frames to be used needs to be considered. When a higher number of frames is used, the frame distance increases for the two frames in which the distance is farthest. It is obvious that the temporal correlation between the two frames decreases as the frame distance increases. For video sequences with complex motion, the decrease in temporal correlation is more evident. In that case, using a small number of frames is better.

2.4 Advantages and Disadvantages

The advantage of using the proposed method is that the temporal redundancy of the adjacent frames can be exploited for intra-frame coding without using techniques for inter-frame coding such as motion estimation. Inter-frame coding consumes the most time during the search for the correlation between adjacent frames. The proposed method enables reducing the computation time for video coding. The disadvantages, however, are the delay and the memory requirement. This proposed method requires an arrangement procedure that is several frames in advance. This causes a delay in the encoding scheme. Also, when the number of frames increases for generating a large picture, a larger amount of memory is needed for both the encoding and decoding. In the decoding procedure, the memory is needed for buffering the large decoded picture, and each frame is rearranged and used as an output. Despite the disadvantages, however, using the proposed method improves the coding efficiency. The temporal scalability is restricted for our proposed method, however, in certain cases using JPEG 2000, it is obtained by using spatial scalability. For example, when four frames are interleaved to form a square (Fig. 3 (2×2)), video sequences with 1/4 frame rate can be extracted by reducing the reconstruction level of DWT by discarding the highest frequency level.

3. Simulation

We conducted a simulation to test the effectiveness of the proposed method. The test video sequences we used were "carphone,""news,""silent " (size: 176×144), "tempete," "paris" (size: 352×288), and "mobile & calendar" (size: 720×576). They were all grayscale sequences. Thirty-two frames were used.

3.1 Correlation Coefficient

Table 1 shows the average spatial correlation coefficient between all frames and the temporal correlation coefficients for the original video sequences. The spatial correlation is the average of the horizontal and vertical correlation. We combined four and sixteen frames in the simulation to generate the large picture, and the generation method was simple, symmetric, and proposed as previously shown in Fig. 2. The arrangement of the frames was set to form a square (2×2, 4×4). Using the proposed method improved the spatial correlation for the video sequences "carphone," "silent," "news," and "paris." However, the spatial correlation coefficients were lower for "tempete" and "mobile & calendar." These two video sequences include shots, which caused more complex movements than the other sequences did. Consequently, the temporal correlation of the adjacent frames was low for these two video sequences, as shown, and the conventional method that was simple and symmetric, provided a higher spatial correlation. A low temporal correlation indicates that the proposed method is less effective and that the spatial correlation also decreases.

3.2 Coding Results

For comparison, the video sequences were encoded by using both the proposed and the conventional method with JPEG-LS [7] and JPEG 2000 [16] codecs. For the JPEG 2000, the number of stages used for a wavelet transform was five, and the size of the code block was 64×64. Table 2 gives the results of the simulation for the lossless coding using JPEG-LS, and Table 3 gives the results of using the JPEG 2000 with five wavelet decomposition levels. In the simulation, we combined four and sixteen frames to generate the large picture. The arrangement of the frames was arranged

Table 1 Comparison of correlation coefficients. (For the spatial correlation, the average horizontal and vertical correlation coefficients of all frames are shown. For the temporal correlation, the average correlation coefficients between the adjacent frames are shown.)

correlation	temporal ρ	spatial$(\rho_H + \rho_V)/2$						
frames	1		4 (2×2)			16 (4×4)		
method	original		simple	symmetric	proposed	simple	symmetric	proposed
carphone	0.987132	0.964733	0.958552	0.967591	0.967500	0.954020	0.967817	0.968526
silent	0.991195	0.953891	0.948107	0.959755	0.966960	0.946610	0.961020	0.969211
news	0.995060	0.913866	0.911551	0.914053	0.952082	0.910171	0.912900	0.964260
paris	0.980016	0.883185	0.881922	0.882744	0.919323	0.907176	0.883050	0.926409
tempete	0.942285	0.897917	0.896161	0.897998	0.889545	0.935874	0.898468	0.844854
mobile	0.942244	0.954351	0.954487	0.954469	0.897176	0.973716	0.954903	0.847985

Table 2 Bitrate comparison of the conventional and the proposed method using JPEG-LS showing an average bitrate of 32 frames [bpp].

codec	JPEG-LS						
method	original	simple	symmetric	proposed	simple	symmetric	proposed
frames	1	4 (2×2)			16 (4×4)		
carphone	3.940637	3.913096	3.911419	3.969934	3.928888	3.893328	4.037415
silent	4.665069	4.631372	4.640013	3.832376	4.590968	4.614327	3.398388
news	4.253374	4.211549	4.200383	3.243243	4.193922	4.164378	2.762202
paris	4.807087	4.792525	4.787405	4.140692	4.563267	4.770747	3.906511
tempete	5.154060	5.145505	5.150124	5.156425	4.900028	5.143525	5.331097
mobile	4.658587	4.651948	4.654501	5.309905	4.381840	4.648169	5.527094
average	4.579802	4.557666	4.557308	4.275429	4.426486	4.539079	4.160451

Table 3 Bitrate comparison of the conventional and proposed method using lossless JPEG 2000 showing an average bitrate of 32 frames [bpp]. (The wavelet decomposition level was five.)

codec	JPEG 2000						
method	original	simple	symmetric	proposed	simple	symmetric	proposed
frames	1	4 (2 × 2)			16 (4 × 4)		
carphone	4.766759	4.748027	4.712151	4.514047	4.751490	4.691604	4.605498
silent	5.175446	5.135604	5.107984	4.603673	5.121508	5.085415	4.220674
news	5.144472	5.103999	5.076902	4.427942	5.096709	5.051166	3.811494
paris	5.481884	5.477645	5.458185	4.965690	5.480392	5.447685	4.726479
tempete	5.649579	5.645777	5.627824	5.340490	5.647520	5.618642	5.571602
mobile	4.864845	4.865966	4.857037	4.861268	4.865948	4.854364	5.365195
average	5.180498	5.162836	5.140014	4.785518	5.160595	5.124813	4.716824

to form a square (2×2, 4×4). For the video sequences "carphone," "silent," "news," and "paris," the proposed method achieved a high compression rate. However, it was less effective with "tempete" and "mobile & calendar." In these two video sequences, the conventional method symmetric gave a higher compression rate. This result has a relationship with the results of the correlation coefficients shown in Table 1. When the correlation coefficient is higher, the compression rate is improved. Figure 4 compares the correlation coefficient and the bitrate for the proposed method. The vertical axis on the left shows the value of the correlation coefficients, and the one on the right shows the bitrate in reverse order. The figure clearly shows that the correlation coefficients are related to the bitrate.

The characteristics of the images vary according to the content. The proposed method allows choosing the number of frames and how to arrange them flexibly. The arrangements are important because they affect the resulting correlation of the generated large picture and the coding efficiency.

First, we considered how to arrange the frames. They should be combined depending on the image characteristics. For example, if the correlation coefficient in the vertical direction is significantly lower than the one in the horizontal direction, combining the frames in the vertical direction is better. Combining the frames in the direction that has a low correlation should increase the effect of the proposed method. In this simulation, the video sequences "news" and "paris" were encoded by using the proposed method with JPEG-LS and JPEG 2000 by combining four frames. The four frames were arranged as follows, horizontally (1 × 4), vertically (4 × 1), and square (2 × 2). The correlation of the original frames and the bitrate are shown in Tables 4 and 5. In both sequences, the horizontal correlation of the original frame is lower than the vertical one. Therefore, the proposed method is more effective with an arrangement in the horizontal direction (1 × 4).

Next, we considered the number of frames to use. When more frames are used, the distance between the two frames that are farthest apart increases. For example, when four frames are combined, the maximum frame distance is three. Obviously, the temporal correlation between two frames decreases as the frame increases; this is even clearer for the video sequences with complex motions. In such

Table 4 Comparison of correlation and bitrate when the arrangement differed for "news" sequence. (Four frames were used.)

original correlation	ρ_H	ρ_V	$(\rho_H + \rho_V)/2$
	0.892581	0.935151	0.913866

	correlation $(\rho_H + \rho_V)/2$	bitrate[bpp] JPEG-LS	bitrate[bpp] JPEG 2000
1 × 4	0.952432	3.113853	4.317235
4 × 1	0.934750	3.294340	4.351454
2 × 2	0.952082	3.243243	4.427942

Table 5 Comparison of correlation and bitrate when the arrangement differed for "paris" sequence. (Four frames were used.)

original correlation	ρ_H	ρ_V	$(\rho_H + \rho_V)/2$
	0.864239	0.902132	0.883185

	correlation $(\rho_H + \rho_V)/2$	bitrate[bpp] JPEG-LS	bitrate[bpp] JPEG 2000
1 × 4	0.923235	4.017968	4.912218
4 × 1	0.907693	4.105200	4.929843
2 × 2	0.919323	4.140692	4.965690

cases, when the number of the frames that are combined increases, the spatial correlation of the generated picture decreases, and the final bitrate of the result deteriorates. In this simulation, the video sequences "silent" and "tempete" were encoded by using the proposed method with the following number of frames: 1 (conventional), 2 (1 × 2, 2 × 1), 4 (1 × 4, 4 × 1, 2 × 2), 8 (2 × 4, 4 × 2), and 16 (4 × 4), as previously shown in Fig. 3. The results are shown in Fig. 5. The vertical axis on the left shows the value of the correlation coefficients, and the one on the right shows the bitrate in reverse order. For the "silent" sequence, as the number of frames increased, the correlation coefficient increased and the bitrate decreased. Thus, the coding efficiency was improved. In contrast, for the "tempete" sequence, using four frames (4 × 1) made for the best result. This is because the "tempete" sequence included complex motions, and therefore, increasing the number of frames was not effective.

3.3 Relation between Correlation Coefficient and Coding Result

Figures 4 and 5 show that there is a relationship between the correlation coefficient and coding efficiency. When the correlation coefficients increase, the compression rate is improved. The effect of the proposed method can be

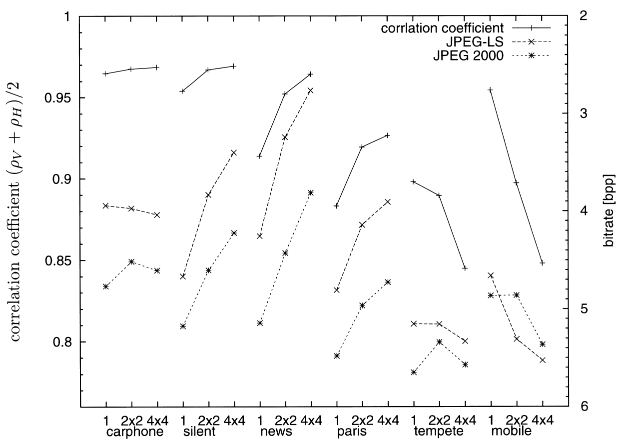

Fig. 4 Comparison of correlation coefficients and bitrate. (The y-axis on the left shows the correlation coefficients and the one on the right shows the bitrate[bpp] in reverse order.)

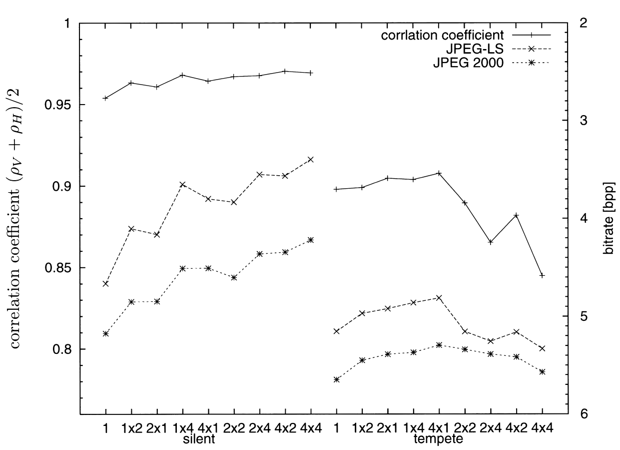

Fig. 5 Comparison of correlation coefficients and bitrate with a different number of frames. (The video sequences were "silent" and "tempete." The y-axis on the left shows the correlation coefficients and the one on the right shows the bitrate[bpp] in reverse order.)

estimated by calculating the correlation coefficients in advance, and the bitrate varies depending on the arrangement of the frames. The proposed method is ineffective on frames where the temporal correlation is low, for example, frames in which the scene change occurs. In such cases, one must judge whether to use the proposed method or not. Correlation coefficients measure the amount of improvement on coding efficiency, and they can be used to control the use of the proposed method adaptively.

4. Conclusion

We described a method for coding video sequences efficiently using an existing lossless encoder. Multiple frames are combined to form one large picture. Doing so improves the coding efficiency. In the proposed method, a temporal correlation is transformed into a spatial correlation. It is effective for video sequences with less movements, as we have demonstrated by coding video sequences using JPEG-LS and JPEG 2000. For the future work, we will compare the coding performance between our proposed method and methods using motion compensation.

Acknowledgement

This work was supported in part by Grant-in-Aid for JSPS Fellows No.2334 from the Japan Society for the Promotion of Science.

References

[1] N. Ono, S. Ishibashi, and N. Kobayashi, "Development of a lossless video coding system using motion compensation," IEICE Technical Report, IE97-140, 1998.
[2] J. Nakajima, Y. Yashima, and N. Kobayashi, "A study on layered lossless video coding using MPEG-2 parameters," IEICE Technical Report, IE99-124, 2000.
[3] N. Ono and Y. Yashima, "An efficiency evaluation of lossless video coding using motion compensation," IEICE Technical Report, IE95-18, May 1995.
[4] G.C.K. Abhayaratne and D.M. Monro, "Embedded to lossless coding 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, Jan. 2001.
[5] T. Nakachi, T. Sawabe, T. Fujii, and T. Fujii, "Multiresolution lossless video coding using inter/intra frame adaptive prediction," IEICE Trans. Fundamentals, vol.E85-A, no.8, pp.1822–1830, Aug. 2002.
[6] D. Brunello, G. Calvagno, G.A. Mian, and R. Rinaldo, "Lossless video coding using optimal 3D prediction," Proc. ICIP 2002, vol.I, pp.89–92, 2002.
[7] ISO/IEC 14495-1, "Lossless and near-lossless compression of continuous-tone still images: Baseline," ISO/IEC, Dec. 1999.
[8] ISO/IEC 15444-1, "JPEG 2000 image coding system, Part 1: Core coding system," ISO/IEC, Dec. 2000.
[9] T. Nakachi, T. Fujii, and J. Suzuki, "Pel adaptive predictive coding based on image segmentation for lossless compression," IEICE Trans. Fundamentals, vol.E82-A, no.6, pp.1037–1046, June 1999.
[10] K. Komatsu and K. Sezaki, "Design of lossless block transforms and filter banks for image coding," IEICE Trans. Fundamentals, vol.E82-A, no.8, pp.1656–1664, Aug. 1999.
[11] M. Iwahashi, S. Chokchaitam, N. Buabthong, P. Zavarsky, and N. Kambayashi, "Lossless scalable coding of images via lossless multichannel prediction," IEICE Trans. Fundamentals, vol.E83-A, no.7, pp.1450–1457, July 2000.
[12] ISO/IEC JTC 1/SC 29 WG 1 N1987: "Motion JPEG 2000 Committee Draft 1.0," ISO/IEC JTC 1/SC 29 WG 1, Dec. 2000.
[13] B. Furht, J. Greenberg, and R. Westwater, Motion Estimation Algorithms for Video Compression, Kluwer Academic Publishers, 1997.
[14] A. Takagi and H. Kiya, "QCIF Video Coding Based on JPEG 2000 Using Symmetry of Images," IEICE Trans. Fundamentals, vol.E85-A, no.8, pp.1898–1901, Aug. 2002.
[15] I. Pitas, Digital Image Processing Algorithms, Prentice-Hall, 1993.
[16] ISO/IEC JTC 1/SC 29/WG 1 WG1N1894, "JPEG 2000 Verification Model 8.6," 2000.