International Journal of Information Technology, Vol. 11, No. 7, 2005

An Efficient Motion-Compensated 3D Wavelet Video Coding System

Zheng Chang[1], Li Zhuo[2], and Lansun Shen[2]

Signal & Information Processing Lab,
Beijing University of Technology, Beijing 100022, China
[1b]changzheng@emails.bjut.edu.cn
[2] {zhuoli, slx}@bjut.edu.cn

Abstract

An efficient motion-compensated 3D wavelet video coding system is proposed in this paper. In this new system, spatio-temporal subbands are generated by MC temporal analysis based on an adaptive GOP structure and 2D spatial wavelet transform, and then encoded by a quadtree-based wavelet image coding method. The experimental results show that, the proposed system can achieve better performance than other ones, while reducing computational complexity and memory requirement to a certain extent.

Keyword: Motion-compensated, wavelet, scalable video coding.

1 Introduction

Scalable video coding (SVC) is widely considered as a promising technology, which allows partial decoding at a variety of resolution and quality levels from a single compressed bitstream[1]. Subband extension of H.264/AVC[2] is one of the SVC methods, which can provide temporal scalability, spatial scalability and quality (SNR) scalability, but at the cost of too much memory requirement and computational complexity. Therefore a practical SVC technology is necessary to be proposed.

Recently, image processing based on wavelet has achieved great success and promoted the research of wavelet video coding technology. Wavelet coding possess the feature of multi-resolution in nature, and no need to construct layered coding structure to realize a variety of scalability. Thus, scalable video coding based on wavelet transform has become a hotspot in the field of scalable video coding.

In this paper, a new wavelet video coding system is proposed. The next section describes the proposed system in general and then analyzes three key techniques in detail. Section 3 provides experimental results. It shows that compared with other coding method[3], the proposed system in this paper is more efficient.

2 Motion-compensated 3D Wavelet Video Coding System

2.1 Block Diagram of the Proposed Video Coding System

The proposed motion-compensated 3D wavelet video coding system is shown in Fig. 1, which includes two key techniques: MC temporal analysis and 3D spatio-temporal subband coding. MC temporal analysis is applied to remove the temporal redundancy and the 2D spatial wavelet analysis the spatial redundancy. The resulted spatio-temporal subbands are encoded by Wavelet Quadtree-based Coding (WQC), and finally form the bitstream with motion vectors.

Similar with MPEG, GOP structure is used in this paper. A GOP can contain 2, 4, 8 or 16 continuous frames. To improve the coding performance, it is necessary to select an appropriate GOP structure according to the motion degree of the video sequences, which is another key technique of the coding system.

The three key techniques of the coding system are analyzed in detail below.

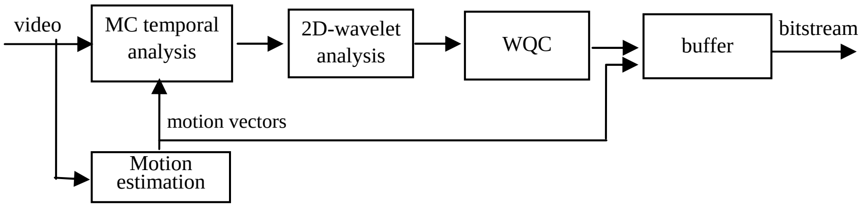
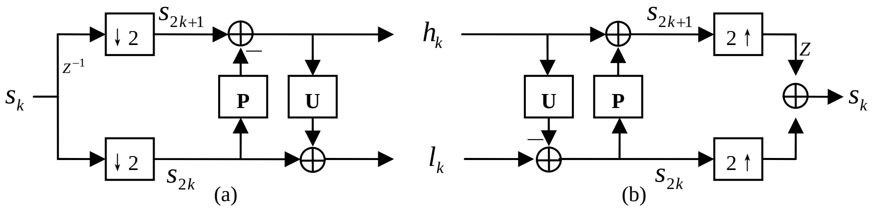


Fig. 1. Block diagram of the proposed video coding system

2.2 MC Temporal Analysis

Because of the high similarity between adjacent frames, efficient video compression significantly relies on effective removal of temporal redundancy in the video sequence. MC temporal analysis, used for removing temporal redundancy in this paper, means that the temporal filtering is performed along the motion trajectory, instead of the direct temporal-axis direction of the input video sequence.

MC temporal analysis can be implemented by temporal DWT based on lifting[4]. The generic lifting scheme consists of three step: polyphase decomposition, prediction, and update. Fig. 2 illustrates the lifting representation of an analysis-synthesis filter bank.



(a) Analysis Filterbank

(b) Synthesis Filterbank

Fig. 2. Lifting representation of an analysis-synthesis filter bank

At the analysis side (a), the odd samples $s[2k+1]$ of a given signal s are predicted by the even samples $s[2k]$ using a prediction operator $\mathbf{P}(s[2k+1])$ and a high pass signal $h[k]$ is formed by the prediction residuals. A corresponding low-pass signal $l[k]$ is obtained by adding the prediction of the residuals $h[k]$ to the even samples $s[2k]$ of the input signal s using the update operator $\mathbf{U}(s[2k])$:

$$h[k] = s[2k+1] - \mathbf{P}(s[2k+1]) \tag{1}$$

$$l[k] = s[2k] + \mathbf{U}(s[2k]) \tag{2}$$

Since both the prediction and the update step are fully invertible, the corresponding transform can be interpreted as critically perfect reconstruction filter bank. The synthesis filter bank simply consists of the application of the prediction and update operators in reverse order.

Let $s[\mathbf{x}, k]$ be a video signal with the spatial coordinate $\mathbf{x} = (x, y)^T$ and the temporal coordinate k . For the extension to MC temporal analysis, the prediction and update operators using the lifting representation of the Haar wavelet are given by

$$\mathbf{P}_{Haar}(s[\mathbf{x},2k+1]) = s[\mathbf{x} + \mathbf{m_{p0}}, 2k] \tag{3}$$

$$\mathbf{U}_{Haar}(s[\mathbf{x},2k]) = \frac{1}{2} h[\mathbf{x} + \mathbf{m_{u0}}, k] \tag{4}$$

where $\mathbf{m}$ is motion vector that can be obtained by any motion estimation method.

After MC temporal analysis, a pair of temporal low- and high-subbands is generated for each two continuous input frames. Low-subband contains the average information of two frames, and high-subband, which contains less energy, can be considered as Displace Frame Difference (DFD) after the motion compensation of two input frames. Therefore, MC temporal analysis can remove temporal redundancy efficiently and thus improve the coding efficiency.

2.3 Adaptive GOP Structure According to Motion Characteristics of the Video Sequences

For the video sequences with low motion, because there is strong correlation among the adjacent frames, large GOP structure should be selected to efficiently remove the temporal redundancy to achieve high performance. But for the video sequences with high motion, small GOP structure should be adopted, which will not affect the coding performance, what is more, it can reduce the memory requirement and computational complexity.

Therefore, in this paper, an adaptive GOP structure according to motion characteristics of the video sequences is proposed. As Fig. 3 shows, four kinds of GOP structure can be chosen, including 2 frames, 4 frames, 8 frames and 16 frames, and labelled as GOP＝2，GOP＝4，GOP＝8 and GOP＝16, respectively.

Table 1 illustrates the times of MC temporal analysis in a variety of GOP structure.

Motion vector is an useful parameter for representing motion characteristics of the video sequences. The formula can be described as follows:

$$\mathrm{MA} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [mv_x^2(i,j) + mv_y^2(i,j)] \tag{5}$$

where [$mv_x(i,j)$, $mv_y(i,j)$] is the motion vector of the pixel (i,j) , and M, N are width and height of the image, respectively. Smaller MA means low-motion of video sequence, thus the large GOP should be selected in this case. Otherwise, the small

GOP should be chosen.

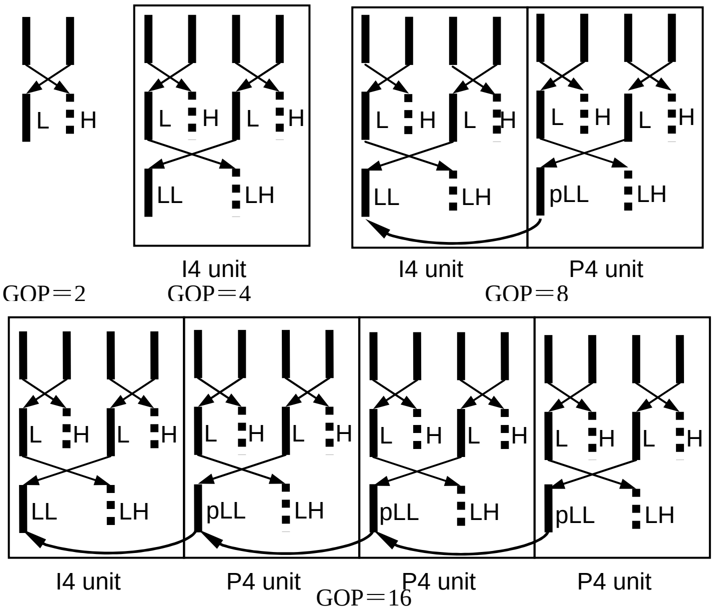


Fig. 3. Four kinds of GOP structure

Table 1. Times of MC temporal analysis in a variety of GOP structure

GOP	Each GOP	Continuous 16 frame
2 frames	1	8
4 frames	3	12
8 frames	7	14
16 frames	15	15

2.4 WQC

The development of wavelet video coding depends on the development of wavelet still image coding. Whether the wavelet still image coding sufficiently utilize the characteristics of the wavelet transformed image or not, has direct influence on the performance of video coding. WQC is proposed to encode the spatio-temporal subbands, which exploits energy clustering in frequency and in space. [5]

Quadtree split is a simple but efficient method for organizing wavelet coefficient. All pixels in a square block form a Set (S), which can be denoted as $\{c_{i,j}\}$, where (i, j) is the position of wavelet coefficient in a wavelet image. For a value n, whether S is significant or not, can be defined as follows: the S is significant if $\max_{(i,j)\in S}\{|c_{i,j}|\} \ge 2^n$, otherwise insignificant. Thus, it can be described as follows:

$$F(n,S)=\begin{cases}1, & 2^n \le \max\limits_{(i,j)\in S}\{|c_{i,j}|\} < 2^{n+1} \\ 0, & \text{others}\end{cases} \tag{6}$$

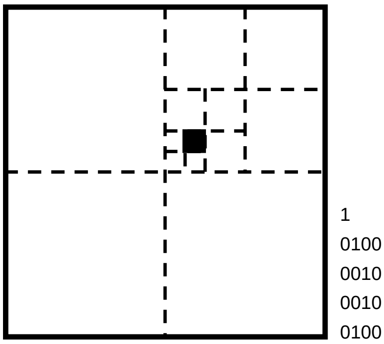


Fig. 4. Quardtree split process

All pixels in the insignificant S can be denoted as 0, and the significant S as 1. The significant S will split further into 4 subsets which have same size, and then each of the subset will be judged whether it is a significant set. For each subset, the same method is adopted to split until obtain all significant coefficients. The process of a quadtree split is shown in Fig. 4.

WQC is a method of bitplane coding in essence, which can generated an embedded bitstream. In each bitplane, coding is carried out in special subbands one by one. Thus, the error propagation can be avoided and the error resilience of the bitstream can be improved by adding some head information during the coding process, it is useful for bitstream transmission in an error-prone environment.

3 Experimental Results

To estimate the efficiency of this system, we compare the proposed system (named as A_MC_3DWQC) with MC_3DSBC proposed in literature [3]. The comparison results at Rc=1.2Mbps is shown in Table 2. Standard test video sequences such as Akiyo, Football, Foreman and Mobile at CIF resolution (96 frames, 30frames/s) were used.

Table 2. Comparison of A_MC_3DWQC with MC_3DSBC for test video sequence (dB)

video sequence	MC_3DSBC	A_MC_3DWQC	gain
Akiyo	40.84	46.21	+5.37
Mother_Daughter	40.24	42.84	+2.60
Football	31.65	33.65	+2.00
Foreman	36.48	37.17	+0.69

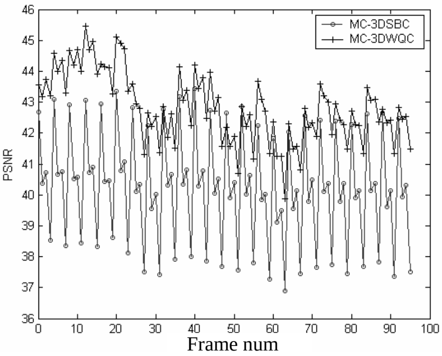


Fig. 5. Frame-by-frame PSNR for the luminance component of the Mother_Daughter

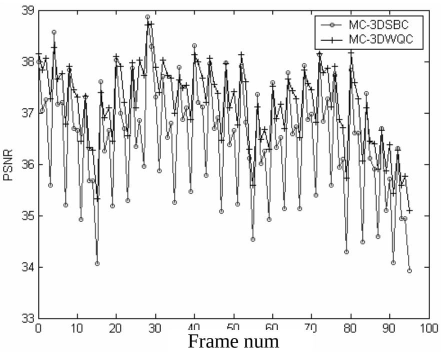


Fig. 6. Frame-by-frame PSNR for the luminance component of the Foreman

Fig. 5 and Fig. 6 illustrate the frame-by-frame PSNR for the luminance component of Mother_Daughter and Foreman, respectively.

It can be seen from the results shown in Table 2, Fig. 5 and Fig. 6 that, under the same coding conditions, our system can achieve better performance for the test video sequence with both low and high motion. The highest coding gain is about 5 dB.

Table 3. Comparison of the adopted GOP structure of the two methods for Football sequence

GOP	MC_3DSBC	A_MC_3DWQC
GOP＝2	0	42
GOP＝4	0	1
GOP＝8	0	1
GOP＝16	6	0

Table 3 tabulates the comparing results of the adopted GOP structure in two systems for Football sequence (96 frames). Because of the high motion characteristics of Football sequence, the structure of GOP=2 was chosen 42 times in A_MC_3DWQC. On the other hand, the fixed GOP=16 structure is adopted in MC_3DSB. So the times of MC temporal analysis in A_MC_3DWQC is 1×42＋3×1＋7×1＝52, but the times in MC_3DSBC is 15×6＝90, which demonstrates that the coding system proposed in this paper can not only improve the coding efficiency, but also reduce computational complexity for the high-motion video sequences.

4 Conclusion

A wavelet video coding method is proposed in this paper. The experimental results show that, compared with MC_3DSBC proposed in literature [3], under the same conditions, for low-motion video sequences, it can get better performance, and for high-motion video sequences, it can not only improve the compression efficiency, but also reduce memory requirement and computational complexity.

Acknowledge

The research work was supported by the National Natural Science Foundation of China (90304001, 60472036), Beijing Natural Science Foundation (4052007, 4032008) and Beijing Municipal Education Commission Project (KM200410005022, KZ200310005004).

References

[1] Shih-Ta Hsiang: Highly Scalable Subband/Wavelet Image and Video Coding [Ph D Dissertation]. Rensselaer Polytechnic Institute (2002).

[2] Julien Reichel, Mathias Wien, Heiko Schwarz: Scalable Video Model 3.0. ISO/IEC JTC1/SC29/ WG11, N6716, October 2004.

[3] Choi Seung-Jong and Woods J W: Motion-Compensated 3-D Subband Coding of Video. IEEE Trans. Image Processing, Vol. 8 (2) (1999), pp. 155-167.

[4] Andrew Secker, David Taubman: Lifting-Based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression. IEEE Trans. Image Processing, ,Vol. 12 (12) (2003), pp. 1530-1542.

[5] Zhuoli: Research on Wireless Image/Video Coding and Transmission Technique [Ph D Dissertation]. Beijing University of Technology (2004).

Zheng Chang, received the Bachelor degree in Electronic Engineering from Beijing University of Technology in 2003. He is a graduate student for Master degree in Signal & Information Processing also in Beijing University of Technology. His research interest includes scalable image/video coding and transmission.

Li Zhuo, graduated from the University of Electronic Science and Technology, Chengdu in 1992. Received the master degree in Signal & Information Processing from southeast University in 1998, and the PH. degree in Pattern Reorganization and Intellectual System from Beijing University of Technology in 2004 respectively. She is an associate professor and the supervisor of postgraduates. She has published over 30 research papers and authored 3 books. Her research interest includes image/video coding and transmission over error-prone network environments.

Lanusn Shen, graduated from Beijing Post and Telecommunication Institute. He is a professor and supervisor of PhD candidate of Beijing University of Technology. He has published over 300 papers and authored over 10 books. His research interest includes image/video coding and transmission, image/video processing in the compressed domain, biomedical image processing, information processing system, etc.