Enhancing the Visual Quality in Hybrid Filters Wavelet-Based Low Bit-Rate Video Codec

Mostafa A. Ahmad, Ayoub Al-Hamadi, Gerald Krell, and B. Michaelis
Institute for Electronics, Signal Processing and Communications (IESK),
Otto-von-Guericke-University Magdeburg
D-39016 Magdeburg, P.O. Box 4210 Germany

mostafa.ahmad@et.uni-magdeburg.de

Abstract

In this paper, we present a hybrid filters wavelet-based video coding technique. The wavelet transformation and subband quantization are developed and optimized in order to enhance the visual quality especially at very low bit-rate. For fulfil the demand of the robustness and flexibility by choosing the optimal frame predictor, several motion estimation methods are examined with inter-frame video coding. The proposed coding technique reduces the bit rate while performs well at average PSNR. In comparison with the conventional coding methods, our proposed coding technique provides significant performance gains in objective and visual quality.

Keywords

Video Compression, Hybrid Video Coding, Discrete Wavelet Transform (DWT), JPEG-2000, Hybrid Wavelet Filters Decomposition.

1 Introduction

Video compression is a rapidly developing technology that is proving useful in a wide range of applications ranging from deep space imaging, medical and security applications to numerous high level task of computer vision. In our modern culture, the financial mechanism of the entertainment industry drives many of our video technology solution. For instance, compressed video enables a variety of applications including video storage on DVD's and Video-CD's, video broadcast over digital cable, satellite and over-the-air terrestrial digital television (DTV), as well as video streaming over best-effort packet networks. Accordingly, very low bit-rate video coding has received considerable attention in academia and industry in terms of both coding algorithms and standards activities. The recently adopted ISO/IEC MPEG and ITU-T VCEG Recommendation H.264 AVC provide a solution for very low bit-rate video applications [1]. Similar to previous video coding standards (such as MPEG1, MPEG2, MPEG4, H.261, H.263, and H.26L), H.264/AVC utilizes transform coding of the prediction residual by using Block Discrete Cosine Transform (BDCT) [1]. Although, BDCT-based coding techniques are widely used for image data compression and have been adopted in many international image and video compression standards, such techniques produce noticeable blocking artifacts along block boundaries in decompressed images at a low-bit-rate. This is because the DCT coefficients in each block of an image are processed and quantized independently. Consequently, an efficient blocking artifacts reduction scheme is essential for preserving the visual quality of decompressed images.

A variety of deblocking schemes for reducing the blocking artifacts have already been proposed to improve the visual quality of block-based coded images at the decoder side [2]-[4]. A wavelet transform-based method is one of the effective techniques that reduce such artifacts [5], [6]. Unfortunately, the wavelet transform-based techniques may suffer from ringing artifacts at very low bit rate compression ratios [7]. The choice of filter length, the depth of decomposition as well as the frame prediction method in wavelet video coder has an important effect on both video quality and compression ratio.

The current paper introduces a coder that better enhance visual quality by allowing different filters with different lengths (hybrid filters) at each stage of the decomposition. The structure of our improved coder is based on JPEG-2000 image compression standard [8]. In intra-frame video coding, a like JPEG-2000 with hybrid filters and varying quantizer step sizes is used to compress each frame independently. In inter-frame video coding, on the other hand, frame difference displacement is used as a simple predictor to exploit the temporal redundancy. At the search about the best frame predictor other motion estimation methods are compared. Exhaustive search, three-step search [9], simple and efficient search [12], and adaptive rood pattern search [13] are compared with the frame difference displacement. The effect of varying the wavelet decomposition levels, adopting the quantizer step size for each subband, choosing the wavelet filter at each stage of decomposition process, as well as choosing the frame prediction method are investigated through an extensive simulation.

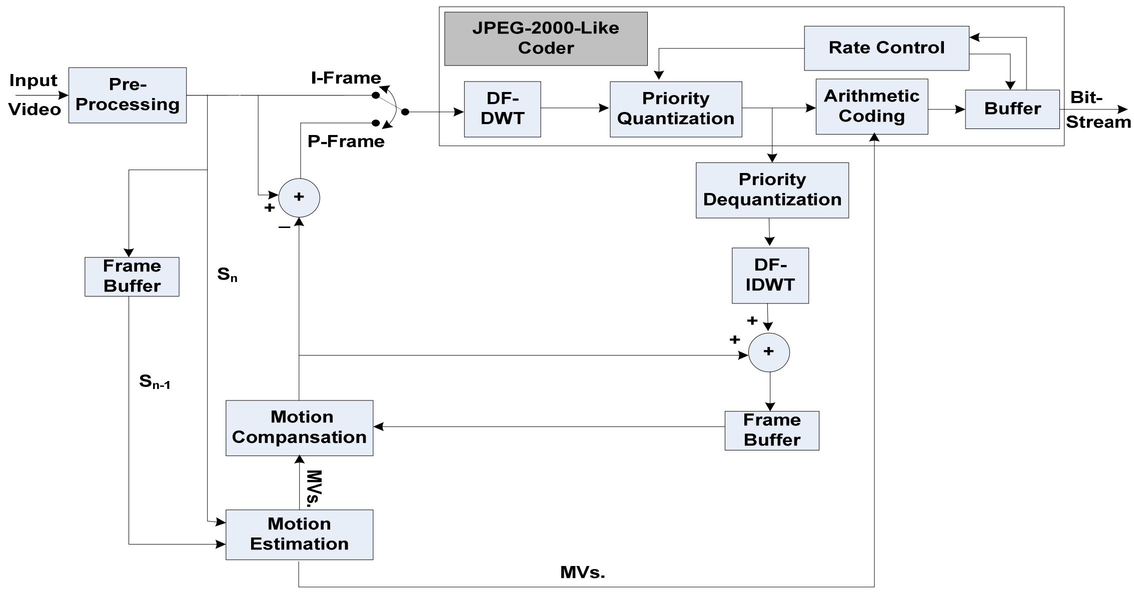


Fig. 1 Wavelet-based Video Codec

The rest of this discourse is organized as follows. In Section 2, the general structure of the coder is described. In Section 3 the idea of using different filters through the decomposition process is explained. In Section 4 the results obtained when coding the test video sequences are presented. The conclusion and future work are addressed in Section 5.

2 Video Codec Description

The general structure of the wavelet-based video codec is shown in Fig. 1. The key processes associated with the video codec can be identified and explained as follows.

The input frame is adjusted to have a nominal dynamic range that is symmetrically distributed about zero. Then, the frame's colors components are transformed by means of irreversible color transform (ICT) to reduce the correlation between them allow more efficient coding [9]. The wavelet transform performs decomposition of video frames into a multiresolution subband representation (i.e., LL, LH, HL, and HH). The wavelet transform scheme is described in the next section in details.

Transform coefficients are quantized using scalar quantization with a deadzone. A different quantizer is employed for the coefficients of each subband, and each quantizer has only one parameter, namely, its step size. We identify a priority scheme to assign the quantizer step size for each subband. The LL subband takes priority over other subbands by halving its step size. The quantizer step sizes are decreased from level to level along the quantization process. That is the first level of subbands have higher quantizer step size than the second level, and the second have higher step size than the third, and so on.

The encoder side of video codec estimates the motion in the current frame with respect to a previous frame. A motion compensated image for the current frame is then created. The motion vectors used for motion compensation are transmitted, as well as the residual frame after discrete wavelet transformed and quantized. The quantized coefficients are then dequantized and inverse discrete wavelet transformed at the encoder to be used as a reference frame for the subsequent frames. Once the residual frame is discrete wavelet transformed, each subband is independently quantized and coded using a bit-plane coder. Arithmetic coding technique is used as an entropy coder [11]. The quantizer step sizes are adjusted in order to control the rate. As the step sizes are increased, the rate decreases, at the cost of greater distortion. Although this rate control mechanism is conceptually simple, it does have one potential drawback. Every time the quantizer step sizes are changed, the quantizer indices change, and the frame encoding must be performed again. Since the frame coding requires a considerable amount of computation, this approach of rate control may not be practical in computationally constrained encoders.

The decoder simulates the processes of dequantization and inverse DWT at the encoder after entropy decoding the bit stream. The decoder reverses the effects of preprocessing in the encoder and creates a full frame.

3 Proposed Approach in DWT

In this section, we present our hybrid filters discrete wavelet transform (HFDWT). The discrete wavelet transform decomposes the video frames into different subbands with both time and frequency information and facilitate arriving at a high compression ratio. The DWT has been made extremely flexible by allowing explicit specification of parameters such as the number of decomposition levels, the type and coefficients of the used filters at each level of the decomposition. While the default irreversible transform in JPEG-2000 is implemented by means of the Daubechies 9-tap/7-tap filter [10], we choose to carry out the transformation by hybrid filters. In the forward wavelet transformation of a color video frame two different filters can be chosen to transform both of the luminance and chrominance components. To investigate the effect of the filter length on the visual quality of decoded video frame, each of the following schemes is examined along the decomposition process:

(i) All frame components (i.e., Luminance and Chrominance) are decomposed by using the same filter.

(ii) The luminance component is decomposed by one filter and the chrominance components are decomposed by other filter with different length.

(iii) The first level of decomposition of luminance component is decomposed by one filter and the remaining levels of luminance as well as chrominance components are decomposed by other filter with different length.

Joining these three schemes with two different filters with different lengths, different cases can be obtained to apply on the video frame through the transformation stage of video coding. Due to the effect of longer filters to avoid artificial blockiness and shorter filters to reduce the extent of ringing artifacts, we expect that using long and short filters together –as hybrid filters– in transformation process will enhance the visual quality of decoded video. To analyze the effect of different wavelet decomposition levels for visual quality, the different cases are repeated with different levels.

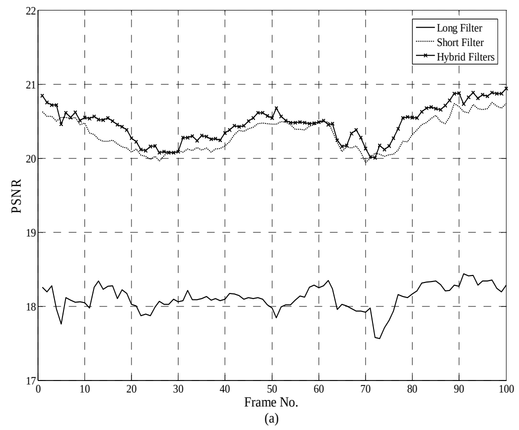
4 Experiments and Results

The proposed video coding system was tested for a variety of raw QCIF colored video sequences for both intra- and inter-frame video coding. The video coder is adopted to encode 10 frames/s at 28 Kb/s. The Daubechies 9-tap/7-tap filter; as an irreversible long filter; as well as the Harr filter; as short filter; are used in the DWT process.

The most popular measure among image and video community is Peak Signal to Noise Ratio (PSNR). The PSNR is used to measure the difference between the original frame and the decoded frame. The coding process is repeated for intra- and inter-frame video coding at different decomposition levels e.g., 2,3,4,5 and 6 levels. By considering the priority quantization scheme as described later, the video coder will be adopted to use this scheme in the rest of the simulation results.

For intra-Frame video the proposed schemes in section 3 are applied. Results shown in Fig. 2a indicate that the best PSNR of average 20.4507 db is for using hybrid filters. That is starting the wavelet decomposition by a long filter for the first level of luminance component and by a short filter for the remaining levels as well as the chrominance components. The nearest competitor is to use a short filter for all luminance and chrominance components. This is due to the effect of longer filter at the start of decomposition to avoid any artificial blockiness. Shorter filter is used at the later levels of the decomposition to reduce the extent of ringing artifacts.

Although, it is acceptable to use the short filter to decompose all frame components, it is disadvantageous to use the long filter that have an average PSNR of 18.1069 dB. The variation of the bit per pixel (bpp) as a function of the frame number is displayed in Fig. 2b. The best scheme to give higher PSNR doesn't affect the bpp. However, the average bpp for the best PSNR scheme (Hybrid Filters) and the bior4.4 are the same of about 0.11 bpp.



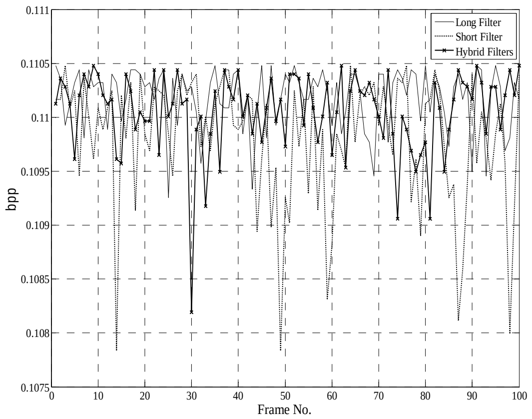
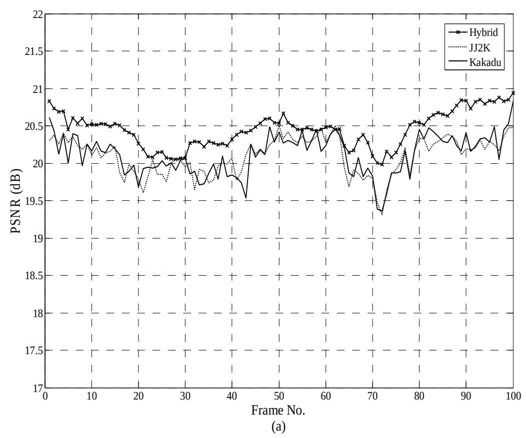


Fig. 2 Foreman QCIF video sequence Intra-Frame coded at 28kbps, 10fps, with 6 decomposition levels, with long, short, and hybrid filters: (a) PSNR and (b) bpp.



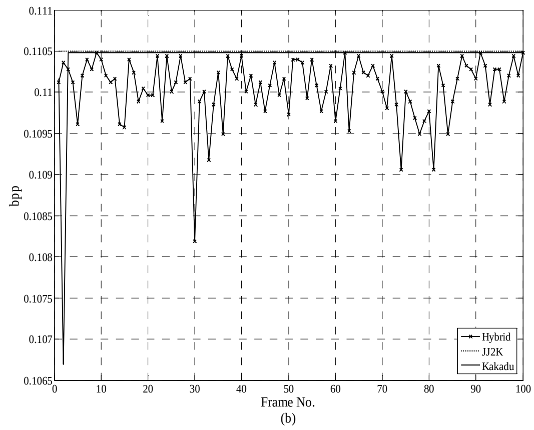
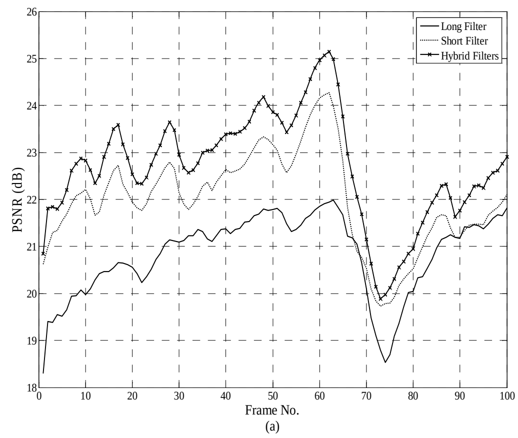


Fig. 3 Foreman QCIF video sequence Intra-Frame coded at 28kbps, 10fps, with 6 decomposition levels, with Kakadu, JJ2K, and Hybrid filters codec: (a) PSNR and (b) bpp.



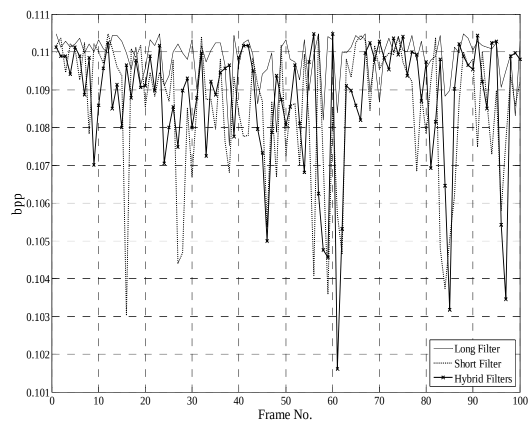
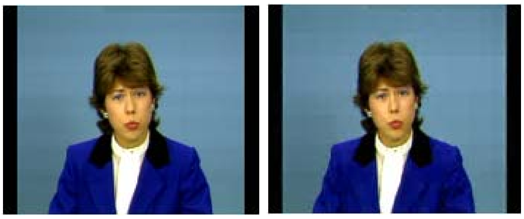


Fig. 4 Foreman QCIF video sequence Inter-Frame coded at 28kbps, 10fps, with 6 decomposition levels, with frame difference displacement predictor: (a) PSNR, and (b) bpp.

(a) (b)

Fig. 5 Foreman Frame 62, (a) Original, and (b) Decoded at 28kbps, 6 decomposition levels, with Inter-Frame hybrid filters codec, PSNR=25.16, bpp=0.038.



(a) (b)

Fig. 6 Claire Frame 47, (a) Original, and (b) Decoded at 28kbps, 6 decomposition levels, with Inter-Frame hybrid filters codec, PSNR=35.19, bpp=0.038.

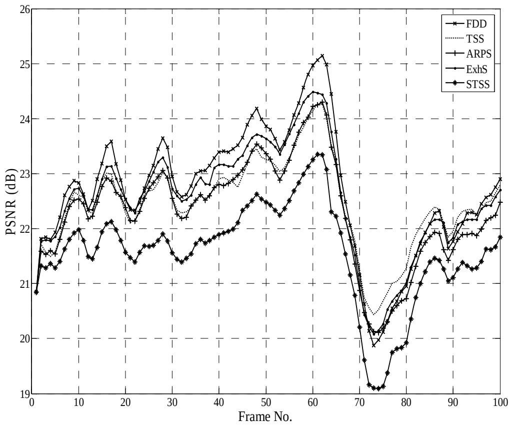


Fig. 7 The PSNR of Foreman QCIF video sequence coded at 28kbps, 10fps, with 6 decomposition levels, and different motion estimation methods.

For comparison our proposed hybrid filters wavelet coder with the recent state-of-the-art wavelet-based image coder, namely JPEG-2000. We use David Taubman's JP2 implementation (Kakadu) that provides free executables [14] as an intra-frame codec. Another important implementation for JPEG-2000 is Java JPEG-2000 (JJ2K) [15]. As shown in Fig. 3a our proposed hybrid filters coder out performs both the Kakadu and

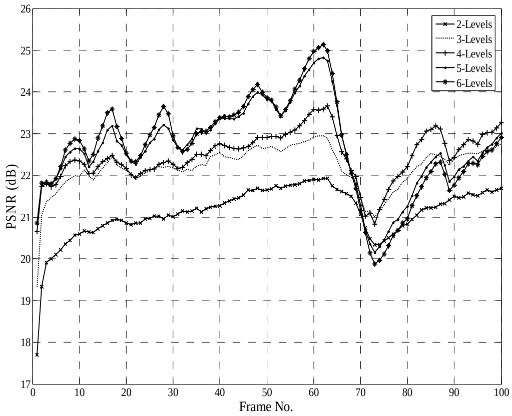


Fig. 8 The PSNR of Foreman QCIF video sequence coded at 28kbps, 10fps, with 6 decomposition levels, and different decomposition levels.

JJ2K in terms of PSNR at the same bit rate. While the hybrid coder has a higher PSNR than Kakadu and JJ2K, it saves the bit per pixel (bpp) under the target bit rate as shown in Fig. 3b.

Results for Inter-frame video coding indicate the same observations as in intra-frame video coding. The best PSNR is achieved at using the hybrid filters (As shown in Fig. 4a). The average gain of PSNR between long and hybrid is 1.9569 dB with 0.0012 increases in average bpp (Fig. 4b). The list of average PSNR and bpp for variety of video sequences are listed in Table 1.

The proposed video codec by the using of two filters in the DWT has a good performance gain in terms of objective and visual quality at a very low bit rate as shown in Figs. 5 and 6. When the decoded frames of Foreman and Claire are compared, you can observe that the Claire frame have a higher PSNR than the Foreman frame. This is due to the smooth motion in Claire video sequence.

Among The various frame predictor methods the Frame Difference Displacement (FDD) has the best PSNR as in Fig. 7. This is because the quality of the predicted frame by Three-Step Search (TSS), Adaptive Rood Pattern Search (ARPS), Exhaustive Search (ExhS) and Simple and Efficient Search (STSS) suffer from the block-based calculation of displacement vectors at the boundary of the frame. Furthermore, the motion vectors add more bits to the bit stream. These overhead bits increase the quantizer step size to match the target bit rate and reduce the quality. The overall average PSNR of FDD, TSS, ARPS, ExhS and STSS for Foreman video are 22.6897, 22.4463, 22.3005, 22.5289 and 21.5594 dB respectively. The all methods share nearly the same bpp.

	Intra-Frame		Inter-Frame	
	Avr. PSNR	**Avr. bpp**	**Avr. PSNR**	**Avr. bpp**
Foreman QCIF Video Sequence				
Long Filter	18.1069	0.1102	22.7360	0.1052
Short Filter	20.3153	0.1098	23.8638	0.1056
Hybrid Filters	20.4507	0.1101	24.6929	0.1064
Claire QCIF Video Sequence				
Long Filter	21.3981	0.1052	24.8857	0.1079
Short Filter	23.3810	0.1089	31.9378	0.1110
Hybrid Filters	23.7882	0.1089	32.8246	0.1125
Carphone QCIF Video Sequence				
Long Filter	19.7121	0.1102	22.8403	0.1094
Short Filter	21.8888	0.1100	24.5373	0.1085
Hybrid Filters	22.0586	0.1101	25.5954	0.1084
Hall-Monitor QCIF Video Sequence				
Long Filter	19.0950	0.1101	21.2397	0.1088
Short Filter	20.4354	0.1098	23.9010	0.1056
Hybrid Filters	20.9135	0.1101	24.4512	0.1061
Mobile QCIF Video Sequence				
Long Filter	13.4693	0.1103	14.9023	0.1100
Short Filter	15.1343	0.1101	16.6101	0.1089
Hybrid Filters	15.2619	0.1101	17.0755	0.1092
News QCIF Video Sequence				
Long Filter	17.6428	0.1102	20.8935	0.1080
Short Filter	19.4128	0.1101	23.9872	0.1045
Hybrid Filters	19.9070	0.1100	24.8795	0.1039
Salesman QCIF Video Sequence				
Long Filter	21.0801	0.1099	22.6608	0.1057
Short Filter	22.3197	0.1099	26.0256	0.1030
Hybrid Filters	22.6511	0.1101	26.6293	0.1025

Table 1 Average PSNR and bpp for both intra- and inter-frame Color QCIF Video sequences.

This is because the quality of the predicted frame by Three-Step Search (TSS), Adaptive Rood Pattern Search (ARPS), Exhaustive Search (ExhS) and Simple and Efficient Search (STSS) suffer from the block-based calculation of displacement vectors at the boundary of the frame. Furthermore, the motion vectors add more bits to the bit stream. These overhead bits increase the quantizer step size to match the target bit rate and reduce the quality. The overall average PSNR of FDD, TSS, ARPS, ExhS and STSS for Foreman video are 22.6897, 22.4463, 22.3005, 22.5289 and 21.5594 dB respectively. The all methods share nearly the same bpp.

Although the overall average gain of FDD is not high in Foreman video sequence, it is good for Claire video at the coast of about 0.0046 bpp. The overall average PSNR of FDD, TSS, ARPS, ExhS and STSS for Claire video are 32.8246, 29.3994, 29.5259, 29.1053 and 27.5698 dB respectively.

Fig. 8 shows that 6-levels of decomposition give better PSNR than the other levels. The overall average gain of 6-levels is of 1.5688, 0.4988, 0.2184, and 0.0203 dB higher than that of the 2-, 3-, 4-, and 5-levels, respectively. This clearly means that visual quality is better in 6-levels. On the other hand, the average bpp of 6-levels is greater than the average bpp of 2-levels by 0.001. The main result of our simulations is that we can adaptively set the number of levels in decomposition process to get more visual quality than any other number of levels for the same compression.

To verify whether higher decomposition levels give higher PSNR, another video sequences are examined such as Hall Monitor and Car Phone. The 5-levels of decomposition gives the best PSNR. It can be concluded that 5- or 6-levels can refine the visual quality of inter-frame video coding - by our improved coder- without nearly any cost in bpp.

5 Conclusion and Future Work

In this discourse, a wavelet video codec is proposed with a new strategy in both the quantization and transformation methods. The effect of giving the LL subband a priority over the other subbands in quantization process, as well as decreasing the quantizer setp size along the decomposition levels on the visual quality is investigated. Also, the effect of using hybrid filters along the decomposition on the quality enhancement in the decoded frames is addressed.

The use of long filter in the first level of the decomposition process with short filter in remaining levels can improve the visual quality in terms of PSNR, with inferior increase in the bpp. While using the hybrid filters can improve the visual quality, the number of decomposition levels has a refinement effect that enhances the visual quality. Apply the priority scheme at the quantization stage add more enhancements to the visual quality. The use of FDD as a simple fast frame predictor and a more suitable method in real-time applications is a good choice especially at the smooth motion video sequence.

Instead of using a supple rate control in the current work, it may be advantageous to employ a more intelligent one suitable for real applications. Furthermore, we may need to investigate the performance of our proposed codec when developing it toward the scalability feature.

6 References

[1] Joint Video Team of ITU-T and ISO/IEC JTC 1, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 ISO/IEC).

[2] G. Qui, "MLP for adaptive postprocessing block-coded images", IEEE Trans. Circuits Syst. Video Technol., vol. 10, no. 8, pp. 1450-1454, Dec. 2000.

[3] H. Paek, R. C. Kim, and S. U. Lee, "A DCT-based spatially adaptive post-processing technique to reduce the blocking artifacts in transform coded images", IEEE Trans. Circuits Syst. Video Technol., vol. 10, no. 1, pp 36-41, Feb. 2000.

[4] P. List et al., "Adaptive deblocking filter", IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 614-619, July. 2003.

[5] Z. Xiong, M. T. Orchard, and Y. Q. Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet represent-tations", IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 2, pp. 433-437, Apr. 1997.

[6] N. C. Kim, I. H. Jang, D. H. Kim, and W. H. Hong, "Reduction of blocking Artifact in block-coded images using wavelet transform", IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 3, pp. 253-257, June 1998.

[7] Shen-Chuan Tai, Chuen-Ching Wang, Ling-Shiou Huang And Ying-Ru Chen, "Morphological De-ringing Filter Design for JPEG-2000", Journal Of Information Science And Engineering 19, 667-680, 2003.

[8] ISO/IEC, ISO/IEC 15444-1, Information technology - JPEG 2000 image coding system, 2000.

[9] Iain E. Richardson, Video Codec Design Developing Image and Video Compression Systems, UK, John Wiley & Sons Ltd, April 2003, Ch. 6,2, pp. 102, 13.

[10] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using the wavelet transform", IEEE Trans. Image Processing, pp. 205-220, Apr. 1992.

[11] Sayood, Khalid, Introduction to Data Compression, San Francisco, Morgan Kaufmann, 2000, Ch. 4, pp. 77-97.

[12] Jianhua Lu, and Ming L. Liou, "A Simple and Efficent Search Algorithm for Block-Matching Motion stimation", IEEE Trans. Circuits And Systems For Video Technology, vol. 7, no. 2, pp. 429-433, April 1997.

[13] Yao Nie, and Kai-Kuang Ma, "Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation", IEEE Trans. Image Processing, vol. 11, no. 12, pp. 1442-1448, December 2002.

[14] Kakadu Software: A Comprehensive Framework for JPEG2000. [Online]. Available: http://www.kakadusoftware.com/

[15] JJ2000: An implementation of the JPEG2000 standard in Java. [Online]. Avaliable: http://jj2000.epfl.ch/