# Improved Performance of Very Low Bit-rate Video Coding Using Wavelet Packet Transform

San Ratansanya, Thumrongrat Amornraksa and Bundit Tipakorn
Department of Computer Engineering, Faculty of Engineering,
King Mongkut's University of Technology Thonburi,
91 Pracha-Uthit Road, Bangkok 10140, Thailand
Tel. +66-2-470-9083, Fax.: +66-2-872-5050
e-mail : hameroon@hotmail.com, t.amornraksa@cpe.eng.kmutt.ac.th

**Abstract:** This paper proposes the use of wavelet packet transform in a transform based video coding scheme, which is mainly used in low/very low bit-rate video coding schemes i.e. H.263 standard. In the experiments, the discrete cosine transform in the video coding scheme is replaced by the wavelet packet transform, and the improved performance in term of peak signal to noise ratio is measured and compared with the results obtained from the coding scheme implementing the ordinary wavelet transform. The experimental results show an impressive improvement obtained from the use of wavelet packet transform.

## 1. Introduction

Video coding is a technique used to compress a large size of digital video data to be as small as possible in order to make use of limited bandwidth of today's networks. Transform based coding is one of the most wisely used techniques in digital image and video coding, and has become a part of standard algorithms used in the real world. Usually, the discrete cosine transform (DCT) is employed because of its fine decorrelation characteristic. H.261 is one of successful low bit-rate video compression standard based on the DCT, which is proposed by ITU-T [1]. The objective of H.261 is to transmit video signal at n×64 kbps over ISDN network. Later in 1996, ITU-T approved H.263 as a new standard [2], while H.263 is in fact the improvement of H.261. It is developed for transmitting the video signal over a very low bit-rate transmission system, i.e. less than 64 kbps. The differences between H.261 and H.263 can be found in [3]. MPEG-4 is another well-known high performance standard for video [4]. It has been developed using the same basic idea as used in H.263 and is being used over many communication networks such as Internet nowadays.

In recent years, much of the research activities in image and video coding have been focused on wavelet-based technique e.g. the discrete wavelet transform (DWT) and the embedded zerotree wavelet (EZW), many of them have achieved a significant improvement in term of image quality. Marpe and Cyon [5] used wavelet-based technique instead of DCT-based, which results in 1-2 dB of PSNR increasing at the same bit-rate. Moreover, this technique can be applied to any video coding standard such as H.263 and MPEG-4. Pao and Sun [6] presented a Sliding-Window rate control technique, which analyze statistically the characteristics of the previous image frames and use the obtained knowledge to compress the current image frame. In [7], the concept of a content-based video coding method was proposed. Since human face is the most important part, compared to the others, the number of bits to represent the face area should be greatest. This method may not achieve a lower output bit-rate, but the quality of coded images was better in total. In 1998, H.263 version 2, also known as H.263+, was officially approved as a standard [8], where some new optional modes were added. More efficient optional modes have been recently developed for the new generation standard, H.263++. However, both of them trade additional complexity for more compression gain.

In this paper, wavelet packet transform (WPT), a class of DWT, is considered to be used in transform coding process in low bit-rate video coding. Since the wavelet packet transform provides more energy compaction than ordinary wavelet transform, the video coding scheme employing this approach can provide an improved performance, in term of peak signal to noise ratio (PSNR), from the decoded video sequence. In other words, with the same quality of decoded video sequence, the encoding scheme employing the WPT can provide lower output bit-rate than that of the DWT. In next Section, the brief introduction to transform based video coding is presented. The fundamental concepts of wavelet and wavelet packet transforms are given in Section 3, while the experimental setting is described in Section 4. The experimental results are presented, and discussions are given in Section 5. The conclusions are finally drawn in Section 6.

## 2. Overview of Low Bit-rate Video Coding

Since this research work focuses on low bit-rate video coding, H.263 based video coding is considered and used in our experiments. As mentioned earlier, the main objective of H.263 video coding standard is to provide the encoded video sequence at low or very low output bit-rate for mobile networks and PSTN applications. The basic configuration of H.263 is mainly based on the ITU-T H.261 i.e. a hybrid of interframe prediction to make use of temporal redundancy, and transform coding of the residual signal to reduce spatial redundancy. Temporal prediction is based on block-based motion estimation and motion compensation, the DCT is used for spatial redundancy reduction. The DCT is also used to transform blocks of 8×8 pixels of motion compensated prediction error. The DCT coefficients are then quantized, and rearranged into one-dimensional array, by scanning them in zigzag order. Finally, motion vectors, and quantization parameters are coded using entropy coding scheme. The block diagram of the H.263 encoding scheme is illustrated in Figure 1.

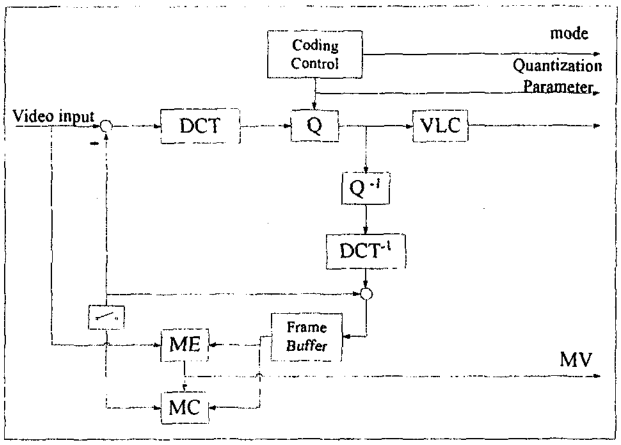


**Figure 1. Block diagram of H.263 video encoder**

Note that baseline H.263 has some improvement over the H.261, such as half pixel motion compensation, new Variable Length Coding (VLC) tables, and some other minor enhancements. Also, H.263 codec scheme supports four optional modes, which are defined in many associated Annexes. The description of these modes can be found in [2]. H.263+ and H.263++ can be considered in the same way as baseline H.263. That is, they support a wide range of custom source format, such as higher refresh rates, resizable computer windows, and wide format pictures, by introducing a number of extra optional modes into the en/decoding scheme. For instance, twelve new optional modes are added to the H.263+ [8], and three more optional modes are added to the H.263++ [9].

## 3. Wavelet and Wavelet Packet Transform

Wavelets are mathematical functions that divide data into different frequency components, and then study each component with a resolution matched to its scale [10]. Each transformation of wavelets involves correlating the wavelet with the given signal (derived from image), where the coefficient values depend on how closely correlated the wavelet is with the given part of the signal. The wavelet is stretched after the entire signal is covered and the process is executed in this fashion repeatedly for all scales. The arrangement of coefficients in the wavelet transform is shown in Figure 2.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 5 | 6 | 17 | 18 | 21 | 22 |
| 3 | 4 | 7 | 8 | 19 | 20 | 23 | 24 |
| 9 | 10 | 13 | 14 | 25 | 26 | 29 | 30 |
| 11 | 12 | 15 | 16 | 27 | 28 | 31 | 32 |
| 33 | 34 | 37 | 38 | 49 | 50 | 53 | 54 |
| 35 | 36 | 39 | 40 | 51 | 52 | 55 | 56 |
| 41 | 42 | 45 | 46 | 57 | 58 | 61 | 62 |
| 43 | 44 | 47 | 48 | 59 | 60 | 63 | 64 |

**Figure 2. Arrangement of coefficients in 8×8 pixels block after taking discrete wavelet transform twice**

Wavelet packets are particular linear combinations of wavelets. They form bases, which maintain many of the orthogonality, smoothness, and localization propeties of their parent wavelets. The coefficients in the linear combinations are computed by a recursive algorithm. The image is then split into sixteen subbands. Let $n$ be a number of iteration of wavelet packet transform. As a result, there have $2^n$ different subbands.

The wavelet packets offer considerably more flexible than the wavelet transforms for dealing with classes of unknown time-frequency characteristics. The arrangement of coefficients in the wavelet packet transforms is shown in Figure 3.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 5 | 6 | 17 | 18 | 21 | 22 |
| 3 | 4 | 7 | 8 | 19 | 20 | 23 | 24 |
| 9 | 10 | 13 | 14 | 25 | 26 | 29 | 30 |
| 11 | 12 | 15 | 16 | 27 | 28 | 31 | 32 |
| 33 | 34 | 37 | 38 | 49 | 50 | 53 | 54 |
| 35 | 36 | 39 | 40 | 51 | 52 | 55 | 56 |
| 41 | 42 | 45 | 46 | 57 | 58 | 61 | 62 |
| 43 | 44 | 47 | 48 | 59 | 60 | 63 | 64 |

**Figure 3. Arrangement of coefficients in 8×8 pixels block after taking wavelet packet transform twice**

## 4. Experimental Settings

In the experiment, we chose four types of video sequence, which are normally used to determine the performance of video coding technique. The first frame of each video sequence (frame no. 0) is illustrated in Figure 4, and its details are summarized in Table 1.

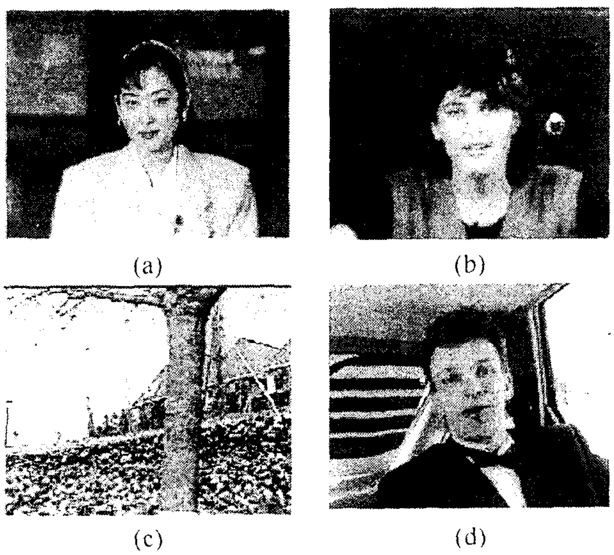


**Figure 4. The first frame of video sequence (a) '*Akiyo*' (b) '*Miss America*' (c) '*Flower*' and (d) '*Carphone*'**

**Table 1. Specifications of the testing video sequences**

| Sequence Name | Frame Number | Frame Format | Video Length (Sec.) |
|---|---|---|---|
| Akiyo | 300 | QCIF 176 x 144 | 10 |
| Miss America | 150 | QCIF 176 x 144 | 5 |
| Flower | 128 | QCIF 176 x 144 | 4.3 |
| Carphone | 382 | QCIF 176 x 144 | 12.7 |

H.263+ in default mode was used for comparative purpose. The experiments were carried out in the steps as shown in Figure 5.

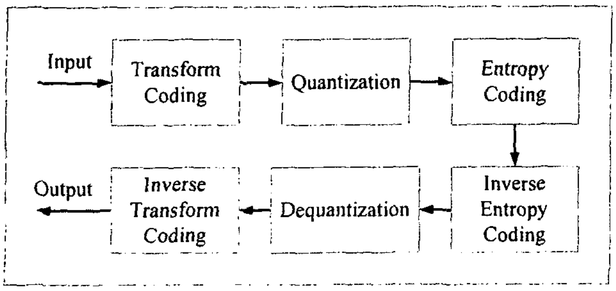


**Figure 5. Block diagram of coding process used in the experiments**

To evaluate the performance of the proposed technique the PSNR is used, and its value can be determined by the following equation.

$$PSNR = 10 \log_{10}\left(\frac{255^2}{\frac{1}{N}\sum_{i=1}^{N}\left(P_{ref}(x,y) - P_{prc}(x,y)\right)^2}\right) \quad (1)$$

where N is the total number of pixels within the image, $P_{ref}(x,y)$ and $P_{prc}(x,y)$ are the pixel values of the reference and processed images, respectively. The summation of PSNR, over the image frames, will then be divided by the total number of frames to obtain the average value. The performance of each technique, the DWT and WPT, is finally compared by using the average PSNR value obtained from the experiments.

## 5. Experimental Results and Discussions

Figure 6 shows the resultant frame no. 4 of the sequence "*Flower*" and "*Carphone*" after being decoded by each technique. After the video sequence was decoded, the average PSNR of each video sequence obtained from two different transformation techniques were measured, and the results are shown in Table 2. Figure 7 illustrates the graphical comparison of the performance between the coding scheme using the DWT and the WPT.

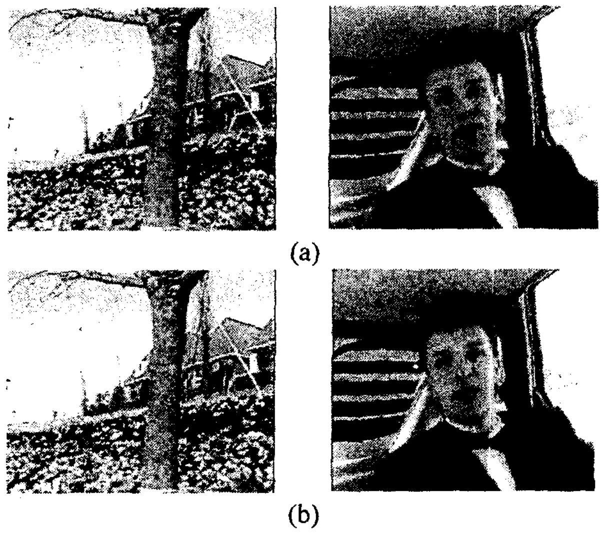


**Figure 6. Samples of resultant frame (frame no. 4) from the coding scheme using (a) DWT (b) WPT**

**Table 2. Performance comparison of the testing video sequences from the coding scheme using DWT and WPT**

| Sequence Name | Average PSNR (dB) | | |
|---|---|---|---|
| | DWT | WPT | Improved |
| Akiyo | 69.84 | 74.58 | + 4.74 |
| Miss America | 70.24 | 74.53 | + 4.29 |
| Flower | 69.25 | 72.79 | + 3.54 |
| Carphone | 67.51 | 71.10 | + 3.59 |
| **Avg.** | **69.21** | **73.25** | **+ 4.04** |

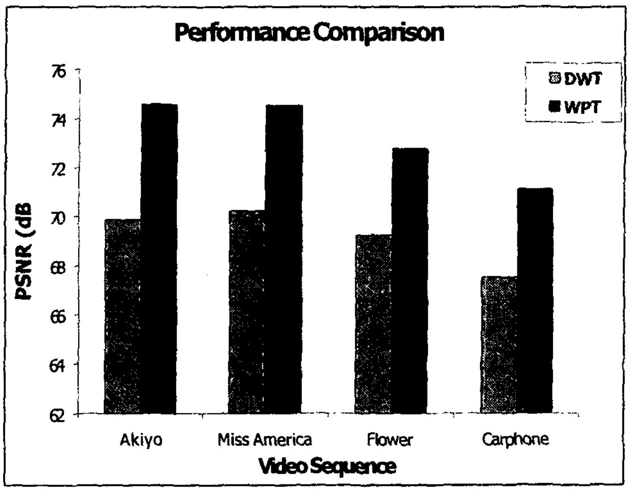


**Figure 7. Performance comparison between the coding scheme using DWT and WPT**

In addition, the processing time required to complete the en/decoding process in the video coding scheme for each technique was measured and compared, the results are shown in Table 3.

**Table 3. Comparison of processing time required in the en/decoding process between DWT and WPT**

| Sequence Name | Processing Time (Sec.) | |
|---|---|---|
| | **DWT** | **WPT** |
| Akiyo | 28.80 | 126.28 |
| Miss America | 30.47 | 127.18 |
| Flower | 30.48 | 142.82 |
| Carphone | 27.88 | 112.76 |
| **Avg.** | **29.41** | **127.26** |

It is obviously seen from Table 2 that the average PSNR obtained from the use of WPT was about 3-4 dB higher than that of the DWT. This is because of the more flexible characteristics of the WPT, where every subband was taken into the decomposing process. Also, the obtained results imply that applying the WPT in the encoding process help improve the performance of the video coding scheme in term of output bit-rate as well.

For the complexity aspect however the processing time required to complete the en/decoding process in the video coding scheme implementing the proposed technique was approximately 4 time larger than that of the scheme implementing the DWT. This is because, in the transform process, the WPT has to perform the analysis and decompose all subbands contained in the 8×8 pixel block, while the DWT need to performed the above process only once on the lowest frequency subband (top-left corner).

## 6. Conclusions

The implementation of the wavelet packet transform in transform based video coding has been proposed in this paper. According to the experimental results, the coding scheme employing the wavelet packet transform provided a better performance in term of PSNR, compared to the scheme employing the ordinary wavelet transform by approximately 4 dB on average. However, the complexity in term of processing time required in the en/decoding process of the wavelet packet transform was about 4 times higher than another one's.


## 7. Acknowledgment

This research work is supported by the Communication Authority of Thailand (CAT) and the National Research Council of Thailand (NRCT).


## References


[1] ITU Telecom. Standardization Sector of ITU, "Video codec for audiovisual services at n×64 kbit/s", *ITU-T Recommendation H.261*, December 1990.

[2] ITU Telecom. Standardization Sector of ITU, "Videc coding for low bit-rate communication", *ITU-T Recommendation H.263*, March 1996.

[3] B. Girod, E. Steinbach, and N. Faerber, "Comparison of the H.261 and H.263 video compression standards", *Standards and Common Interfaces for Video Information Systems*, K.R. Rao, editor, Critical Reviews of Optical Science and Technology, Philadelphia, Pennsylvania, October 1995, Vol. 60, pp. 233-251.

[4] Wu, D., Hou, Y.T., Zhu, W., Zhang, Y.Q. and Chao, H.J., "MPEG4 compressed video over the Internet", *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems (ISCAS '99)*, Vol. 4, 1999, pp. 327 -331.

[5] D. Marpe and H. L. Cycon, "Very low bit-rate video coding using wavelet-based techniques", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, Issue: 1, February 1999, pp. 85-94.

[6] Pao, I.-M. and Ming-Ting Sun, "Encoding stored video for streaming applications", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.11, Issue: 2, Feb. 2001, pp. 199-209.

[7] J. Karlekar and U. B. Desai, "Content based very low bit-rate video coding using wavelet transform", *Proceedings of International Conference on Image Processing (ICIP 99)*, Vol. 2 , 1999, pp. 66-70.

[8] G. Côté, B. Erol, M. Gallant and F. Kossentini, "*H.263+ video coding standard: complexity and performance*", Department of Electrical Engineering, University of Biritish Columbia. Vancouver, Canada.

[9] ITU Telecom. Standardization Sector of ITU, "Draft for H.263++, Annexs U, V and W to Recommendation H.263", *ITU-T Recommendation H.263*, February 2002.

[10] G.S. Burrus, A. R. Gopinath and H. Guo, "*Introduction to Wavelet and Wavelet Transform*" Houston, Rice University, 1996.