Video Coding Using Embedded Wavelet Packet Transform and Motion Compensation

Po-Yuen Cheng and Jin Li and C.-C. Jay Kuo

Signal and Image Processing Institute, Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, CA 90089-2564

ABSTRACT

In this paper, we propose a hybrid video coder using the wavelet packet transform and motion compensation. To reduce the computational complexity of the wavelet packet decomposition, a novel criterion based on the rate-distortion function is proposed. After the wavelet packet transform, the successive quantization scheme and the adaptive arithmetic coder are applied for lossy wavelet coefficient compression. Extensive experiments have been carried to compare the embedded wavelet packet coder with the MPEG standard and the embedded pyramid wavelet coder. The proposed wavelet packet transform method has the best performance in terms of both visual quality and PSNR measurement.

Keywords: video compression, wavelet packets, rate-distortion function, codec model, embedded coding.

1 INTRODUCTION

Video compression is an important topic which has been intensively studied in recent years. A video coder often decomposes the video signal into two parts: (1) the motion compensated prediction which uses the strong temporal correlation between adjacent frames, and (2) the displacement frame difference (DFD) which is the prediction error after motion compensation. Since the compressed DFD usually constitutes a major portion of the entire coded bit stream, effective compression of DFD is critical. In traditional video compression methods such as MPEG [3], the algorithm for compressing DFD is the same as that for compressing the still image frames. They do not take into account that the characteristics of the two types of images are different. That is, a still image frame often contains smooth surfaces while DFD is primarily formed by textured regions so that DFD has significant middle and high frequency components. This special property of DFD can be exploited to achieve efficient compression by using the *wavelet packet* transform.

The use of the wavelet packet transform in signal compression was first examined by Coifman *et al.* [1], and several wavelet decomposition criteria were considered. They included entropy measurement, thresholding and bit counts. Another wavelet packet decomposition criterion based on the rate-distortion tradeoff was recently introduced by Ramchandran and Vetterli [6]. Their basic idea was to determine the best wavelet packet decomposition which achieves the minimum distortion under a certain bit rate constraint. The cost function used is $D + \lambda R$, where D is the distortion measurement, R is the coding bit rate and λ is the optimal operating slope. In their algorithm, λ needs to be determined iteratively. Besides, D and R have to be estimated from coding. Therefore, the computational complexity required by this scheme is very high.

In this research, we also consider the optimal wavelet packet decomposition in the rate-distortion sense, but reduce the computational complexity by considering a rate-distortion coding model for the wavelet coefficients. It is shown that the R-D performance of the quantizer and the entropy coder used in our algorithm can be modeled by an exponential decaying curve. With this model, we derive a R-D optimized wavelet packet decomposition criterion which requires a much lower computational complexity while still providing a good coding performance. Another important feature in the proposed new algorithm is the embedding property of the compressed bit stream. By embedding, we mean a bit stream which is organized in a decreasing importance order so that one can truncate the bit stream at any point while maintaining a perceptible decoded image. A successive quantization scheme is applied to achieve this embedding property. Finally, an adaptive arithmetic coder is used for entropy coding. The quantization method and the entropy coder are similar to those used in the layer zero coding proposed by Taubman and Zakhor [8] with some improvements [4]. Extensive experiments have been performed in comparing the MPEG standard, the embedded pyramid wavelet coder and the embedded wavelet packet coder. The proposed wavelet packet transform method has the best performance in terms of both visual quality and PSNR measurement (about 2-2.5 dB better than MPEG).

This paper is organized as follows. The fast wavelet packet transform with the rate-distortion criterion is described in Section 2. A general framework for wavelet video coding and the properties of DFD are discussed in Section 3. Experimental results are presented in Section 4. Concluding remarks and possible extensions are given in Section 5.

2 FAST WAVELET PACKET TRANSFORM USING RATE-DISTORTION CRITERION

In the wavelet packet transform, a specific decomposition structure which provides the optimal rate-distortion (R-D) tradeoff has to be selected from a large wavelet packet decomposition pool. A straightforward way to obtain the best wavelet packet transform is to actually encode images with different wavelet packet transforms and measure the corresponding R-D performance. However, the computational cost of this approach is too high to be practical. Another method is to compute measures such as the energy compaction factor and the entropy of the wavelet packet bands, etc. and use them for decision. Despite being simple, these measures do not correlate well with the R-D coding performance. In this section, we present a fast wavelet packet decomposition using a model-based R-D criterion. We will first examine the R-D behavior of the quantizer and the entropy coder adopted in our proposed compression scheme, and then use this R-D function in the development of the fast wavelet packet decomposition.

2.1 Rate-Distortion Model for Layer Zero Coding (LZC)

The quantizer and the entropy coder adopted for the coding of wavelet packet coefficients in our proposed compression scheme are very similar to those used in the layered zero coding (LZC) method introduced by Taubman and Zakhor [8] with some minor improvements as described in [4]. In LZC, bits of wavelet coefficients are coded layer-by-layer, i.e. starting from the most significant bit (MSB) of the coefficients and gradually proceeding to the least significant bit (LSB). At each layer, quantization results of the previous layer are refined and bits of each layer are entropy encoded. The quantization and reconstruction rules at bit layer i for wavelet coefficient can be specified as follows:

$$Q_{i,j} = \begin{cases} 0, & \mid W_j \mid < T_i, \\ +1, & W_j \geq T_i, \\ -1, & W_j \leq -T_i, \end{cases} \quad \text{and} \quad \hat{Q}_{i,j} = \begin{cases} 0, & Q_{i,j} = 0 \\ 1.5 \cdot T_i, & Q_{i,j} = 1, \\ -1.5 \cdot T_i, & Q_{i,j} = -1, \end{cases} \tag{1}$$

where $Q_{i,j}$ and $\hat{Q}_{i,j}$ denote, respectively, the quantized and reconstructed bit values for the wavelet coefficient W_j, and T_i is the significant threshold at layer i. Usually, the initial threshold T_1 is chosen to be one half of the maximum absolute value of the wavelet coefficients and $T_{i+1} = T_i/2$ for $i = 1, 2, \cdots$. The quantization procedure can be repeated until the required bit rate is met. The successive quantization converts the wavelet coefficient into a layer representation of bits ordered according its importance. Such a bit layer is then encoded by using the context adaptive arithmetic coder [2]. For details of the coder, we refer to [4] and [8].

In the following, we study the rate-distortion (R-D) behavior of the layered zero coding of a single wavelet coefficient. After the wavelet transform, we assume that the wavelet coefficients in a specific band are identically distributed and their distribution can be modeled by a generalized Gaussian probability density function (PDF):

$$\begin{aligned} P(x) &= ae^{-[b|x|]^\gamma}, \\ a &= \frac{b \cdot \gamma}{2\Gamma(1/\gamma)}, \\ b &= \sigma^{-1}\left[\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}\right]^{1/2}, \end{aligned} \tag{2}$$

where γ is a shape parameter, constant a and b are functions of γ and variance σ^2. Therefore, a generalized Gaussian PDF is completely determined by its shape parameter γ and variance σ^2. Note that both the Gaussian and Laplacian distributions are special cases of the generalized Gaussian distribution with $\gamma = 2$ and 1, respectively. The R-D performances of the layered zero coder (LZC) for a generalized Gaussian distributed source with $\gamma = 0.7$, 1.0 (Laplacian) and 2.0 (Gaussian) are plotted in Fig. 1.

The '+' symbols denote the experimental data points. For comparison, we plot the theoretical Shannon lower bound (SLB) of the source with a dotted line in Fig. 1, where SLB is computed by using the following formula

$$R = h(X) - \frac{1}{2}\log(2\pi e)D, \tag{3}$$

and where $h(X)$ is the entropy of the source X. We see from the figure that the performance of LZC is very close to that of the theoretical SLB at low bit rates. Also, the experimental data can be well approximated with an exponentially decaying model, i.e.

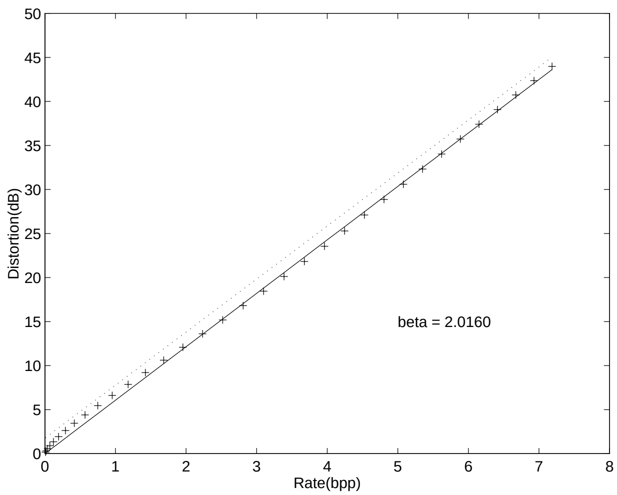
$$D = \sigma^2 2^{-\beta R}, \tag{4}$$

where the optimal parameter β can be obtained by doing the least square fit. The best approximating rate-distortion functions are ploted with solid lines in Figs. 1(a)-(c) with $\beta = 2.0160$, 1.9248 and 1.8694, respectively.

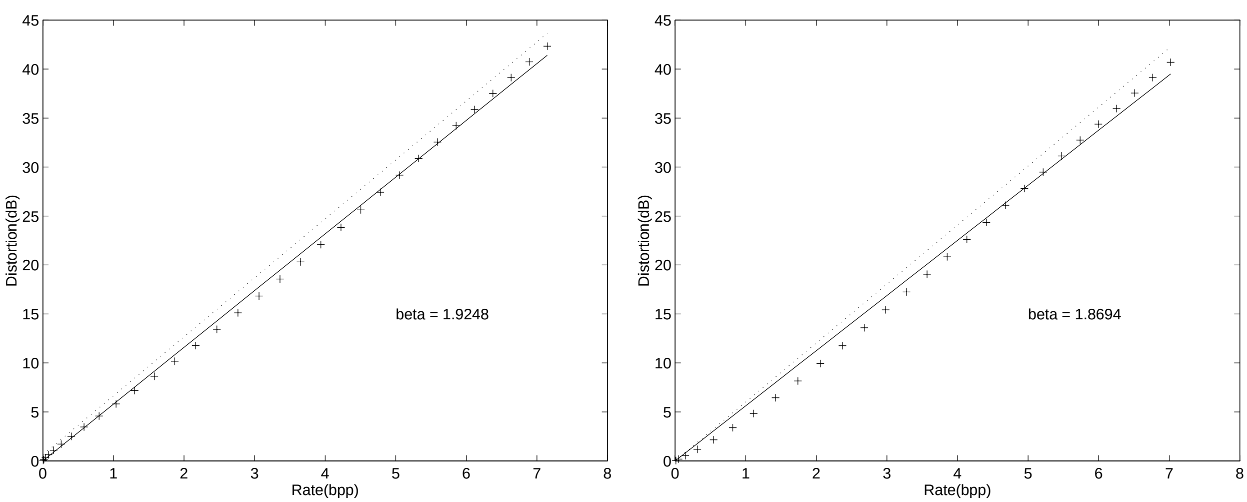
2.2 Wavelet Packet Decomposition

Assume that there are totally N subbands after the wavelet transform. We use $\sigma_1, \cdots, \sigma_N$ to denote the standard deviations and $S_1, \cdots, S_N$ to represent the numbers of wavelet coefficients for these subbands. The total number of coefficients is $S = \sum S_i$. For a total bit budget R, we want to determine the optimal bit rates $R_1, \cdots, R_N$ allocated to these subbands. This can be formulated as the following constrained optimization problem:

$$\begin{cases} \sum S_i R_i \le R, & \text{(the total number of bits is constrained)} \\ D_{avg} = \min \sum S_i D_i, & \text{(the distortion is minimized)} \end{cases} \tag{5}$$



(a)



(b) (c)

Figure 1: The rate-distortion performances of the layered zero coding of three sources: (a) the generalized Gaussian PDF with $\gamma = 0.7$, (b) the Laplacian PDF and (c) the Gaussian PDF. The dotted line corresponds to the Shannon lower bound (SLB) for each source, the '+' symbols are the experimental data points, and the solid line denotes the approximating rate-distortion function with $\beta = 2.0160,\ 1.9248,\ 1.8694$, respectively.

By using the Lagrangian method, the solution of (5) is obtained by

$$\frac{\partial D_i}{\partial R_i} = \text{constant}.$$

Note that

$$\frac{\partial D_i}{\partial R_i} = \sigma_i^2 2^{-\beta_i R_i} \cdot (\beta_i \ln 2) = D_i \cdot (\beta_i \ln 2), \tag{6}$$

where β_i is a R-D efficiency parameter which is only related to the PDF shape parameter γ_i. Since β_i is insensitive to the change of γ_i as shown in Fig. 1, we will consider the following approximation:

$$\beta_i \approx \beta, \quad \text{for} \quad i = 1, \cdots, N. \tag{7}$$

Thus, by using (6) and (7), we can convert the constant R-D slope criterion for optimum bit allocation to the constant distortion criterion. By combining the nonnegative bit rate constraint $R_i \geq 0$, we can derive the approximate solution of (5) as

$$D_i = \begin{cases} D \quad (\text{constant}), & D < \sigma_i^2, \\ \sigma_i^2, & D \geq \sigma_i^2. \end{cases} \tag{8}$$

This implies that to optimally encode a WP decomposed image at a certain bit rate, we only have to determine the corresponding distortion parameter D. For the band with variance σ_i^2 less than D, the coefficients do not have to be encoded at all. For the band with variance σ_i^2 larger than D, we should assign bits to reduce the MSE.

With the above analysis, we can derive the R-D performance with respect to a specific wavelet packet decomposition. Consider that an image is decomposed into N wavelet packet subbands with band size S_i and variance σ_i^2, for $i = 1, \cdots, N$. For a given coding distortion D, the required bit budget $R = \sum S_i R_i$ for the whole image can be computed via

$$R = \frac{1}{\beta} \sum_{i=1}^{N} S_i \log_2 \frac{\kappa_i^2}{D}, \tag{9}$$

where κ_i^2 is called the lower bounded variance of the wavelet packet subband i defined as

$$\kappa_i^2 = \begin{cases} \sigma_i^2, & \sigma_i^2 > D, \\ D, & \sigma_i^2 \leq D. \end{cases} \tag{10}$$

Thus, to satisfy the same distortion requirement, the difference in the bit budget between two different wavelet packet decomposition schemes with total N_1 and N_2 subbands is:

$$\frac{1}{\beta} \left[\sum_{i=1}^{N_1} S_{1,i} \log_2 \kappa_{1,i}^2 - \sum_{i=1}^{N_2} S_{2,i} \log_2 \kappa_{2,i}^2 \right]. \tag{11}$$

Now, let us consider whether to split a a particular subband into 4 smaller subbands. Before and after the split, the band numbers are $N_1 = 1$ and $N_2 = 4$, and the band sizes are:

$$S_{2,1} = S_{2,2} = S_{2,3} = S_{2,4} = \frac{1}{4} S_{1,1}.$$

By using (11), one can compute the bit rate difference as

$$R_s = \frac{S_{1,1}}{\beta}\left[\log_2 \frac{\kappa^2}{[\kappa_1^2 \cdot \kappa_2^2 \cdot \kappa_3^2 \cdot \kappa_4^2]^{1/4}}\right]. \tag{12}$$

If the rate saving is positive, the decomposition is accepted. Otherwise, it is rejected. In other words, we only have to compare the following two values

$$\kappa^2 \text{ and } (\kappa_1^2\kappa_2^2\kappa_3^2\kappa_4^2)^{\frac{1}{4}}$$

for decision making.

To compare the computational complexity of our algorithm with others, we use the pyramid wavelet transform as the reference and compare the execution time of the wavelet packet transform with respect to it. The Lena image with 4-level decomposition was used in the test. The execution time of our proposed wavelet packet decomposition scheme was 1.66 times of that of the pyramid wavelet transform. Compared with the factor of 5.4 for the method described in [7], our method has a much lower complexity.

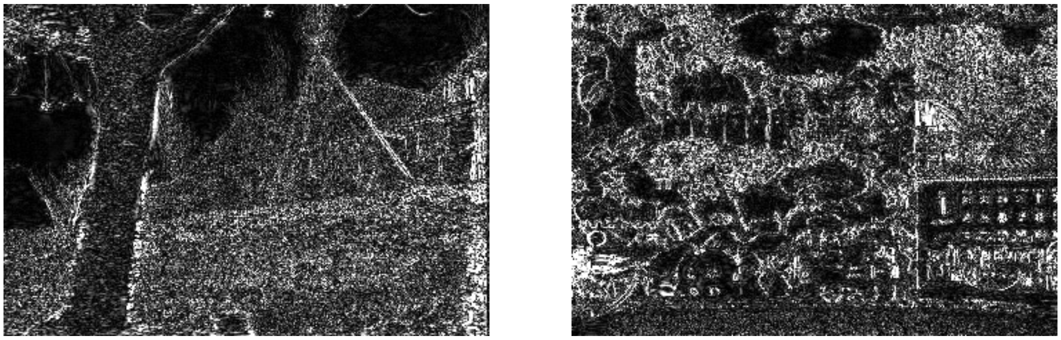
3 VIDEO CODER BASED ON WAVELET PACKET TRANSFORM

3.1 Properties of DFD

In this section, we demonstrate that DFD is more suitable to be encoded using the wavelet packet decomposition. One typical DFD is shown in Fig. 2. To make it visible, we multiply the pixel magnitude by a factor of 6 for display. Visually, one can see that DFD is a texture-like image. This fact can be further supported by the following observations.

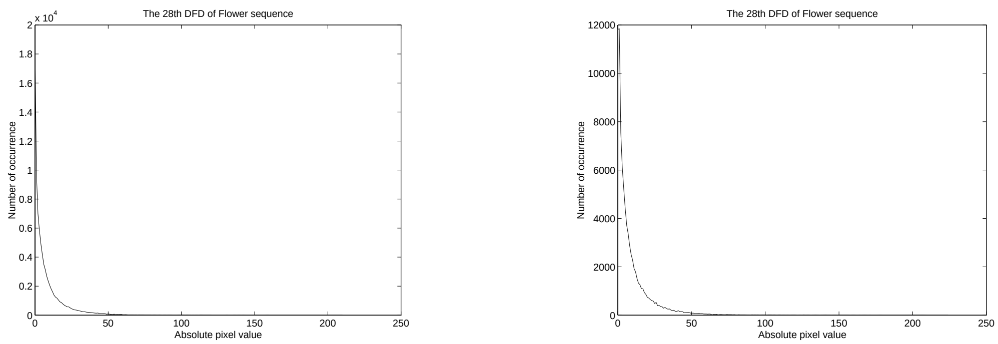
1. DFD has the generalized Gaussian distribution.
   We plot the probability distribution function of DFD obtained with two motion compensation methods, i.e. overlapped block motion compensation (OBMC) and full block motion compensation (FBMC). The result is shown in Figure 3. The distributions can be both well approximated by the generalized Gaussian distribution. This figure also indicates that OBMC outperforms FBMC, since DFD of OBMC has on the average a smaller magnitude than that of FBMC.

2. Energy is more concentrated in the middle and high frequency subbands for DFD than that for the original still images.
   To illustrate this property, we perform 1-level wavelet decomposition on DFD and calculate energy of the 4 subbands. The energy distributions of DFD for several videos are shown in Table 2. We also list the energy distributions of several still images in Table 1. It is clear that DFD cannot be efficiently compacted by using the pyramid wavelet transform. For ordinary images, the pyramid wavelet transform can compact above 90% of energy in the LL subband. In contrast, only about 50% of the total energy is in the LL subband for DFD. Therefore, the wavelet packet transform is more suitable for DFD compression. Actually, based on the energy distributions in Table 2, we can predict whether the wavelet packet DFD coding can outperform the pyramid wavelet transform. For example, for the football sequence, the LL subband of DFD has a higher percentage of energy so that the difference between the wavelet packet transform and the pyramid wavelet is not high. In contrast, for the mobile sequence, the LL subband has a lower percentage of energy so that the coding result of using the wavelet packet should be better. These statements will be verified in the experimental section.



(a) 28th DFD of Flower sequence (b) 30th DFD of Mobile sequence

Figure 2: Typical DFD images of the flower and mobile sequences (enhanced by a factor 6).



(a) DFD with OBMC (b) DFD with FBMC

Figure 3: The distribution of DFD in the spatial domain with different motion compensation methods.

3. DFD is localized in the spatial domain and correlated in the temporal domain.
   Motion compensation often fails in some region due to a complex motion type or occlusion. As a result, DFD is localized in the spatial domain. Also, since the motion field has a strong temporal correlation, DFD is highly correlated along the temporal direction.

3.2 Proposed Video Coder

The flow chart of our proposed algorithm for video compression is shown in Fig. 4. For comparison, we implement the MPEG with the full block motion compensation (FBMC). We also implement a video coder which uses the pyramid wavelet transform to compress the DFD. In the wavelet packet video coder, we first reduce the temporal redundancy by motion analysis (motion estimation and compensation) and then reduce the spatial redundancy by the proposed wavelet packet coder. In motion analysis, we use the exhaustive search for motion estimation and OBMC for motion compensation. Note that exhaustive search is also adopted in the MPEG standard. Thus, motion vectors used in MPEG and the wavelet packet video coder are the same. The main difference is the following. For FBMC, every macroblock is represented by one motion vector and the pixels in the macroblock are only affected by this motion vector. For OBMC, pixels in the macroblock are affected not only by the motion vector of this macroblock but also by the motion vectors in neighboring macroblocks. The prediction gain of OBMC is about 0.2-0.7 dB better than FBMC. For more details of OBMC, we refer to [5]. All coders use only forward motion compensation, (i.e. the first frame is the I frame while the remaining ones are P frames.) In our video coder, DFD is

	Natural image				Texture image			
	Lena	Boat	Barbara	Baboon	Bark	Brick	Bubble	Raffia
LL	99.34%	97.81%	94.13%	86.85%	93.15%	87.37%	94.07%	91.77%
LH	0.48%	1.79%	5.42%	2.71%	3.52%	6.90%	2.66%	3.03%
HL	0.19%	0.26%	0.23%	8.72%	2.72%	3.70%	2.67%	4.89%
HH	0.09%	0.24%	0.22%	1.72%	0.61%	2.03%	0.66%	0.38%

Table 1: Energy distributions in the four subbands for still images after 1-level wavelet decomposition.

	Average energy weight of DFD							
	Flower	Mobile	Football	Tennis	Cheer	Bicycle	Missa	Sales
LL	44.71%	30.60%	76.55%	51.77%	79.64%	61.58%	58.76%	46.22%
LH	12.93%	33.90%	7.06%	23.33%	8.84%	9.18%	20.19%	20.83%
HL	35.79%	18.47%	14.24%	15.54%	9.86%	25.8%	10.12%	25.78%
HH	5.57%	17.03%	2.15%	9.36%	1.67%	3.44%	10.92%	7.16%

Table 2: Energy distributions in the four subbands of DFD after 1-level wavelet decomposition which are obtained by averaging 150 frames of the image sequences.

compressed by using the wavelet packet transform.

4 EXPERIMENTAL RESULTS

In the experiment, we use 8 test image sequences. The test videos are Flower, Mobile, Football, Tennis, Cheer, Bicycle, Miss America, and Salesman. The first six sequences are of size 352×240, and the last two are of size 352×288. We have not yet implemented a rate-control scheme for the proposed video coder. However, since both the pyramid wavelet and the wavelet packet DFD coders are embedded coders, we can easily adjust their coding bit rates to be the same as that of MPEG for each frame. As a result, all video coders in the experiment result in the same bit rate, and their corresponding average PSNR (Peak Signal to Noise Ratio) values are compared in Table 3. In this table, we also show the average bit rate used in coding. In Figs. 5 and 6, The PSNR values for methods with the wavelet packet and the pyramid wavelet transforms are plotted as functions of the frame number. The results indicate that the wavelet packet transform outperforms the pyramid wavelet transform in most cases for about 0.2-0.7dB. The only sequence that the wavelet packet transform performs worse than the pyramid wavelet is the Miss America sequence (0.05 dB worse). The phenomenon can be explained by two reasons. First, DFD of Miss America has a high energy concentration in the LL subband (Table 2) so that the pyramid wavelet has already achieved efficient energy compaction. Motions in the Miss America sequence are smooth and slow and, therefore, they can be well described by motion vectors and DFD are smoother than others. Second, the bit rate in the coding of Miss America is so low that the overhead bits required by the wavelet packet transform can influence the coding result. In image coding, the overhead of wavelet packet is 0.0022 bpp, which is about one tenth of the total coding bit rate (0.024 bpp) in the coding of Miss America sequence. For sequences with more complicated motions (which often implies that motions in these sequences are more difficult to be characterized by the motion vectors), the wavelet packet transform performs better

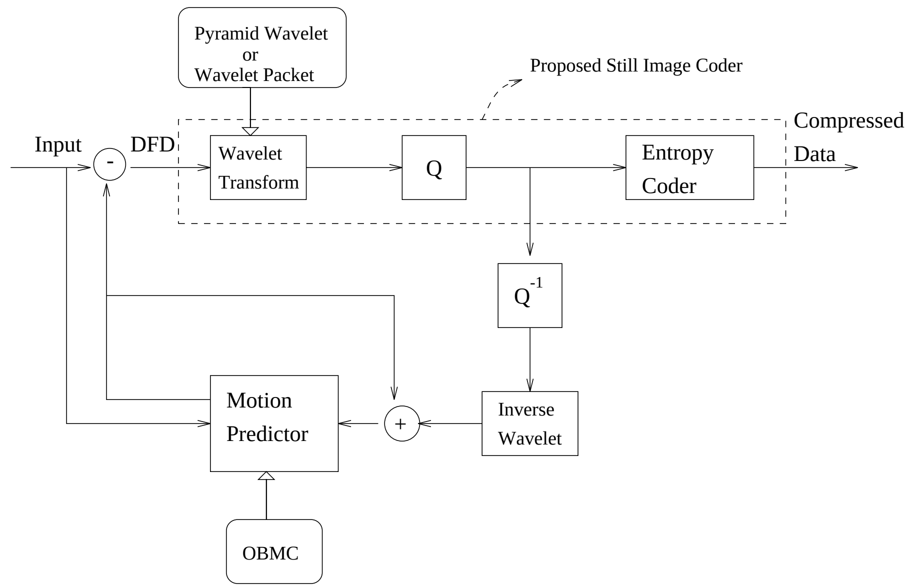


Figure 4: The flow chart of the proposed video coder.

than the pyramid wavelet transform.

5 CONCLUSION

In this work, we presented a fast wavelet packet decomposition algorithm which provides the optimal rate-distortion tradeoff. The resulting wavelet packet transform was then applied to the compression of DFD in video coding where the special characteristics of DFD can be effectively exploited. By combining the wavelet packet transform with the successive quantization scheme and the context adaptive arithmetic coder, the proposed video coder outperforms MPEG by about 2-2.5 dB.

6 REFERENCES

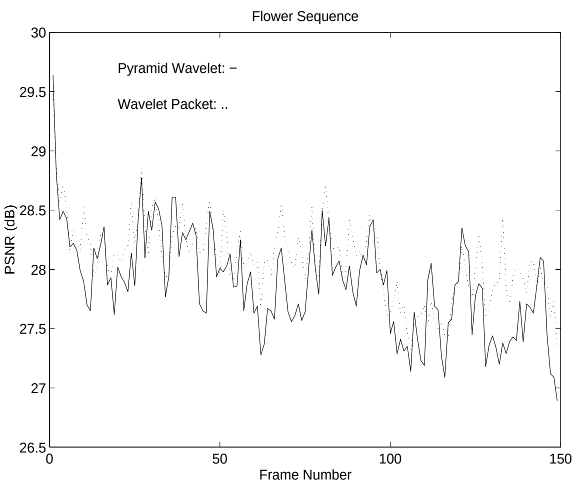
[1] R. Coifman, Y. Meyer, S. Quake, and V. Wickerhauser, "Signal processing and compression with wave packets," in *Reprinted from Numerical Algorithms Research Group, Yale University*, (New Haven, Connecticut), 1990.

[2] ISO/IEC-JTC1/SC2, "Progressive bi-level image compression," ISO Standard CD 11544, ISO, Sep. 1991.

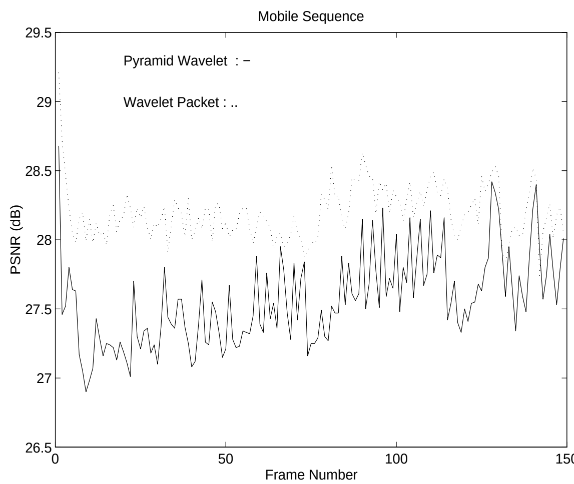
[3] ISO/IEC-JTC1/SC29/WG11, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s," ISO Standard CD 11172, ISO, 1991.

[4] J. Li, P.-Y. Cheng, and C.-C. J. Kuo, "An embedded wavelet packet transform technique for texture compression," in *SPIE's Symposium on Optical Science, Engineering and Instrumentation*, (San Diego, CA), July 1995.

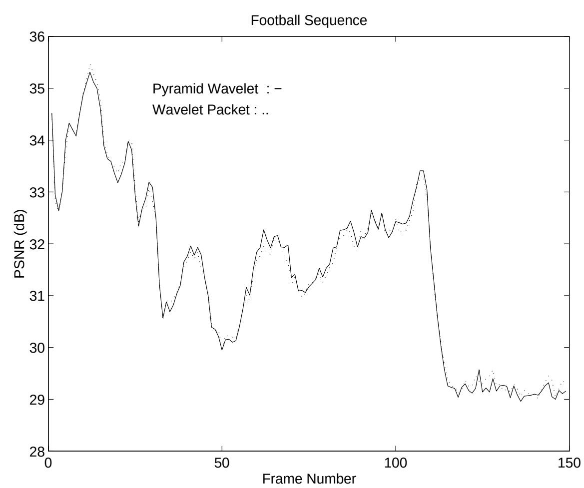
[5] M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: An estimation-theoretic approach," *IEEE Trans. on Image Processing*, No. 5, pp. 693–699, Sep. 1994.



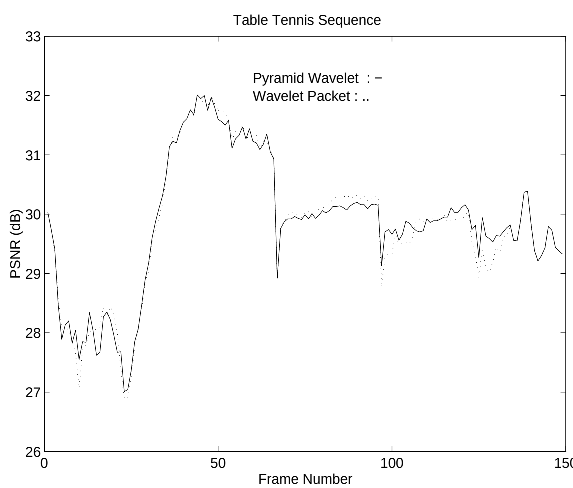
(a) Flower sequence



(b) Mobile sequence

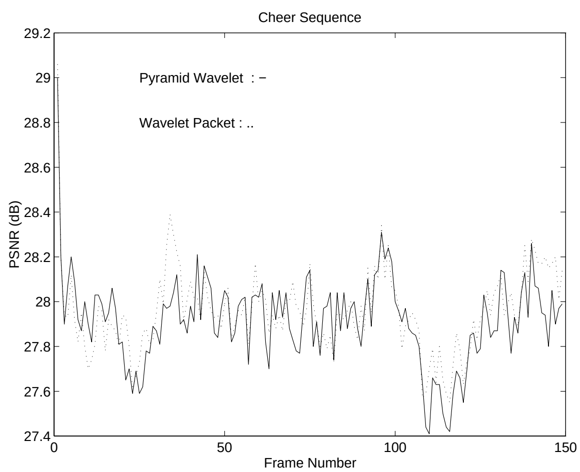


(c) Football sequence

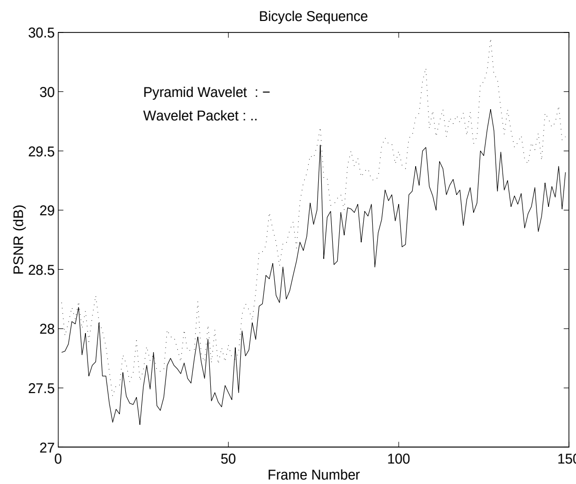


(d) Tennis sequence

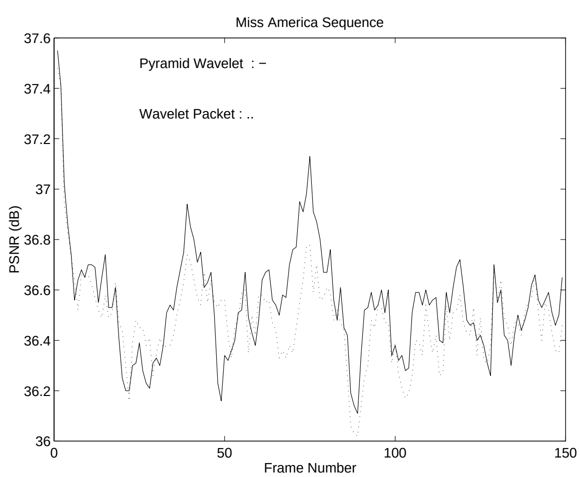
Figure 5: Experimental results for video sequence (I).



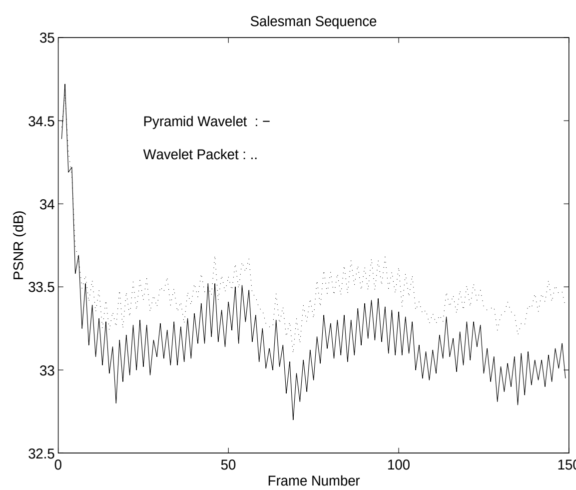
(a) Cheer sequence



(b) Bicycle sequence



(c) Miss America sequence



(d) Salesman sequence

Figure 6: Experimental results for video sequence (II).

	Average PSNR (dB)			Average Bit Rate (bpp)
	MPEG	PY	WP	-
Flower	26.10	27.93	28.12	0.625
Mobile	25.59	27.61	28.23	0.797
Football	29.74	31.54	31.54	0.266
Table Tennis	28.87	29.85	29.86	0.219
Cheer	27.07	27.96	28.35	0.535
Bicycle	26.91	28.49	28.92	0.638
Miss America	34.73	36.56	36.51	0.024
Salesman	31.62	33.19	33.47	0.052

Table 3: Average PSNR and bit rate of the video coding with different methods (PY: pyramid wavelet, WP: wavelet packet).

[6] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. on Image Processing*, No. 2, pp. 160–175, Apr. 1993.

[7] K. Ramchandran, Z. Xiong, K. Asai, and M. Vetterli, "Adaptive transforms for image coding using spatially-varying wavelet packets," *Submitted to the IEEE Trans. on Image Processing*, 1995.

[8] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. on Image Processing*, No. 5, pp. 572–588, Sep. 1994.