

# A VIDEO CODER USING ZEROTREE WAVELET AND HIERARCHICAL FINITE STATE VECTOR QUANTIZATION

İlker KILIÇ

Celal Bayar Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Muradiye/Manisa

Geliş Tarihi : 05.09.2006

## ABSTRACT

A video compression technique with comparable low bit rates is presented in this paper. The overlapping block motion compensation (OBMC) is combined with discrete wavelet transform which followed by Lloyd-Max quantization and zerotree wavelet (ZTW) structure. The novel feature of this coding scheme is the combination of hierarchical finite state vector quantization (HFSVQ) with the ZTW to encode the quantized wavelet coefficients. It is seen that the proposed video encoder (ZTW-HFSVQ) performs better than the MPEG-4 and Zerotree Entropy Coding (ZTE).

**Key Words :** Video compression, Wavelet transform, Zerotree structure, Hierarchical Finite state vector quantization.

## SIFIR AĞAÇ DALGACIK YAPI VE HİYERARŞİK SONLU DURUM VEKTÖR NİCEMLEYİCİ İLE VİDEO KODLAMA

### ÖZET

Bu çalışmada düşük bit oranlarında çalışabilen bir video sıkıştırma tekniği sunulmuştur. Algoritmada, Örtüşmeli Blok Hareket Kompanzasyonu Ayrık Dalgacık Transformü ile birleştirilip ardından Lloyd-Max Nicemleyici ve Sıfır Ağaç Dalgacık Yapısı kullanılmıştır. Bu kodlama tekniğinin en yeni özelliği ise Hiyerarşik Sonlu Durum Vektör Nicemleyicisinin, dalgacık katsayılarını sıkıştırmak için Sıfır Ağaç Dalgacık Yapısı ile birleştirilmiş olmasıdır. Önerilen bu yeni video sıkıştırıcısının MPEG-4 ve Sıfır Ağaç Entropy Kodlama tekniklerinden daha iyi olduğu görülmüştür.

**Anahtar Kelimeler :** Video sıkıştırma, Dalgacık transformü, Sıfır ağaç yapı, Hiyerarşik sonlu durum vektör nicemleyici.

### 1. INTRODUCTION

The proposed video encoder is similar to other motion compensated, block based, discrete cosine transform (DCT) video coding standards such as MPEG-1, MPEG-2, H.261, H.263 and MPEG-4 (Rao and Hwang, 1996 ; Sikora, 1997). In this introduced encoder, the discrete wavelet transform (DWT) is used instead of DCT. The new diamond search (NDS) (Zhu and Ma, 2000) is used as a block motion estimation algorithm to track the local motion. Overlapping block motion compensation (OBMC) (Auyeung et al., 1992) is added to the NDS

to match better DWT, and for coding the wavelet coefficients, the ZTW structure is combined with the HFSVQ. DWT seems to have better performance than DCT for image and video coding (Yeung, 1997). DWT also provides us the scalability functionalities of image and video. The specific components of the coding algorithm as shown in Figure 1 are the NDS block motion estimation algorithm to track the local motion, OBMC to remove the temporal redundancy, DWT to remove the spatial correlation and the Lloyd-Max quantization (Lloyd, 1982) to quantize the wavelet coefficients.

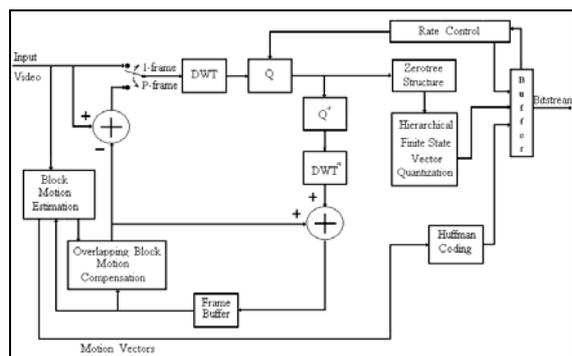


Figure 1. The block diagram of the proposed encoder

The proposed encoder structure is similar to the ZTE scheme introduced in (Martucci et al., 1997). The main difference is combining the ZTW structure with HFSVQ instead of arithmetic coding. It is important to note that, the ZTW structure consists of different size of wavelet blocks that contain similar information. Because of that it is very suitable for these blocks to be coded by HFSVQ algorithm. ZTW structure itself has the ability to control the bit rate. When the HFSVQ is applied to any ZTW structure, then each type of wavelet block will be represented by different size of codeword. That will give us the second opportunity of controlling the bit rate.

The OBMC technique reduces the block artifacts that is used after a block motion estimation algorithm. At high compression ratios, ringing is a major problem of the DWT. The proposed coding algorithm overcomes this problem by allowing different wavelet filter lengths at each stage of the decomposition.

The results of the new introduced video encoder were given for slow motion videos only in (Kiliç and Yılmaz, 2003). In this paper the proposed technique is also simulated for the fast video such as Coastguard and the performance is compared to the standard video compression techniques. It is seen that for both slow and fast videos the PSNR performance of the new introduced encoder is better than the standard techniques.

## 2. PROPOSED ENCODER STRUCTURE

### 2.1. Motion Estimation, Compensation

The proposed video coding algorithm uses the block motion estimation technique called New Diamond

Search (NDS). The NDS performs better than the other well known block based search algorithms such as the three step search, the logarithmic search, the novel four step search (Po and Ma, 1996), the simple and efficient search (Lu and Liou, 1997) and the one at a time search both in reaching the lowest mean square error and achieving a quite small average search point numbers. In block motion estimation algorithm as a performance criteria, the mean square error is used. Motion estimation is performed on the luminance 16x16 blocks. The search area is constrained to 7 pixels in all four directions from the center of the macroblock. The motion vectors of the macroblocks are Huffman encoded.

The OBMC technique is an advanced scheme in H.263 and also used in our proposed coding algorithm. The OBMC technique overlaps, windows, and sums prediction blocks in order to reduce the block artifacts. In this method, each 8x8 block is overlapped with two significant neighbouring blocks:

- 1 The upper left 8x8 block of each 16x16 macroblock is overlapped with the adjacent blocks located at the above and left sides;
- 2 The upper right 8x8 block of each 16 x 16 macroblock is overlapped with the adjacent blocks located at the above and right sides;
- 3 The lower left 8x8 block of each 16x16 macroblock is overlapped with the adjacent blocks located at the below and left sides;
- 4 The lower right 8x8 block of each 16x16 macroblock is overlapped with the adjacent blocks located at the below and right sides. Therefore, each block of motion compensated frame is a weighted sum of three prediction values from the previous reconstructed frame: one value predicted using the current block motion vector and two other values predicted using the neighbouring motion vectors. This OBMC algorithm provides a coherent motion compensated frame, which is free of artificial block discontinuities. Note that intra blocks are not overlapped with their neighbouring blocks.

### 2.2. Discrete Wavelet Transform

The wavelet transform performs decomposition of video frames or motion compensated residuals into a multiresolution subband representation. An

important feature of our implementation is the ability to use different filter banks at each level of the decomposition. This is important because longer filters provide good frequency localization but can cause ringing artifacts along the edges of objects while the use of shorter filter banks cause less ringing artifacts but more blockiness in the reconstructed frame. Therefore the Biorthogonal 9.3 filter bank is used at the first level of the DWT decomposition followed by the Haar filter for the remaining two levels. The scaling function coefficients  $h_s$ , and wavelet function coefficients  $h_w$  of Biorthogonal 9.3 and Haar are given in (1), (2) respectively.

$$\begin{aligned} h_s(1) = h_s(9) &= 0.03314563, h_s(5) = 0.99436891 \\ h_s(2) = h_s(8) &= -0.06629126 \\ h_s(3) = h_s(7) &= -0.17677669 \\ h_s(4) = h_s(6) &= 0.419844651 \end{aligned} \quad (1)$$

$$\begin{aligned} h_w(1) = h_w(3) &= -0.35355339 \\ h_w(2) &= 0.707106781 \end{aligned} \quad (2)$$

### 2. 3. Lloyd-Max Quantization

Lloyd-Max quantization algorithm is a nonuniform quantization process. The distributions of the wavelet coefficients of any video frame is not uniform. Except for the lowest frequency coefficients, the histograms of all remaining discrete wavelet transform layers have a nearly zero symmetric Gaussian distribution that is shown in Figure 2.

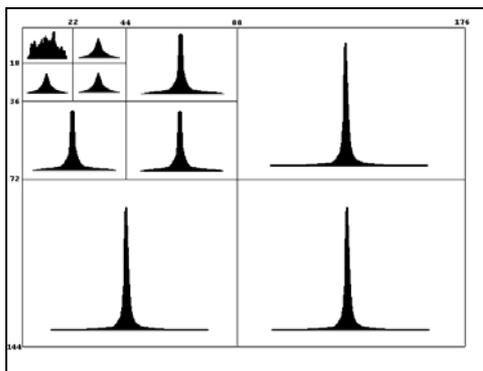


Figure 2. Histogram of three level wavelet transform of a video frame

Lloyd-Max quantization process follows the Gaussian curve and finds an optimum number of nonuniform quantization intervals depending on the standard deviation criteria. In our proposed encoder,

only the coarsest subband is uniformly quantized and the rest of the subbands are nonuniformly quantized by using Lloyd-Max algorithm. At the end of the nonuniform quantization process each standard deviation and the number of quantization intervals for each subband are sent as side information. Note that the Lloyd-Max algorithm is applied to entire wavelet coefficients in a subband except deadzone area in which all amplitudes are too small and set to zero by a predefined threshold value before the process. The application area of Lloyd-Max quantization algorithm is shown in Figure 3.

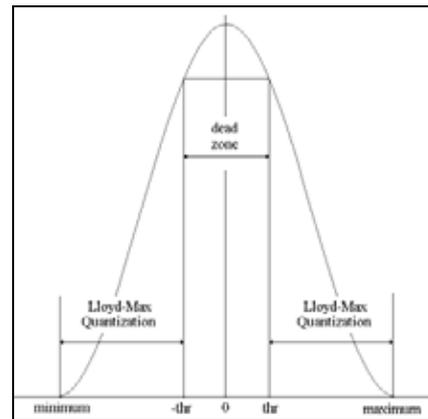


Figure 3. Deadzone and Lloyd-Max quantization locations in a wavelet transform layer. Horizontal scale is the wavelet coefficient amplitudes, vertical scale is the number of wavelet coefficients.

### 2. 4. Zerotree Wavelet Structure

The DWT decomposes the input frame into a set of subbands of varying resolutions. The coarsest subband is a lowpass approximation of the original frame, and the other subbands are finer scale refinements. In this multiresolution subband system, except the highest frequency subbands, every coefficient at a given scale can be related to a set of coefficient of similar orientation at the next finer scale. The coefficient at the coarse scale is called the parent, and all coefficients at the same spatial location and of similar orientation at the next finer scale are called children (Martucci et al., 1997). The parent children relationship is shown in Figure 4.

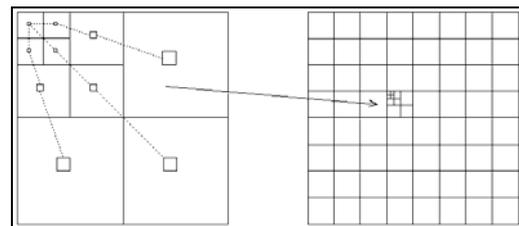


Figure 4. Parent children relationship, reorganization of a wavelet tree into a zerotree block.

In ZTW structure, the coefficients of each wavelet tree are reorganized to form a wavelet block as shown in Figure 4. Each wavelet block comprises those coefficients at all scales and orientations that correspond to the frame at the spatial location of that block. The concept of the wavelet block provides an association between wavelet coefficients and what they represent spatially in the frame. The Lloyd-Max quantization of the DWT coefficients can be done prior to the construction of the wavelet tree, as a separate task, or quantization can be incorporated into the wavelet tree construction.

In the ZTW structure if all coefficients of any parent block are zero or nearly zero then all other children blocks of that zerotree are unimportant. Therefore that zeretree is represented by the value of zero. Since a zerotree collects different size of blocks then it is very suitable for coding those blocks by a variable block size coding algorithm. Therefore in our proposed video coding algorithm, the HFSVQ (Yu, 1994) is used to encode the zerotrees.

## 2. 5. HFSVQ and Rate Control

The HFSVQ is an advanced algorithm of vector quantization that uses different size of blocks in the processing of an image. A zerotree consists of variable block sizes of the similar information that belongs to different layers. The proposed encoder produces the codebooks and codewords of each subband using LBG algorithm (Lin and Tai, 1998). Similar sizes of wavelet blocks are used to produce the codebook list in a DWT layer. The amplitudes and the bandwidths of the wavelet coefficients decrease dramatically from the coarsest low frequency subband to the high frequency subbands. Therefore the size of the codebooks for the low frequency subbands are defined bigger than the codebooks for the higher frequency subbands. Hence, the big children blocks are represented by a few number of bits which increases the coding efficiency.

In the proposed encoder, the standard QCIF video frames are used and three DWT are applied to each video or error frame. Therefore three types of wavelet blocks are used by a zeretree structure, which are 2x2, 4x4 and 8x8 pixels sizes.

Since each QCIF error frame contains a few number of significant wavelet coefficients, then small sizes of codebooks are defined for each DWT level. The number of codeword requirements for each DWT layer of an error frame are, 4-16 for layer 1, 8-32 for layer 2 and 16-32 for layer 3. The intra frame coding requires more codewords than the error frames do. The number of codeword requirements for each

DWT level of an intra frame is, 8-32 for layer 1, 16-64 for layer 2 and 32-128 for layer 3. Here, the layer 1 represents the wavelet coefficients of the highest frequency and the layer 3 represents the wavelet coefficients of the lowest frequency. The HFSVQ coding tree and the block sizes are shown in Figure 5. All the codebooks belong to DWT layers are Huffman encoded.

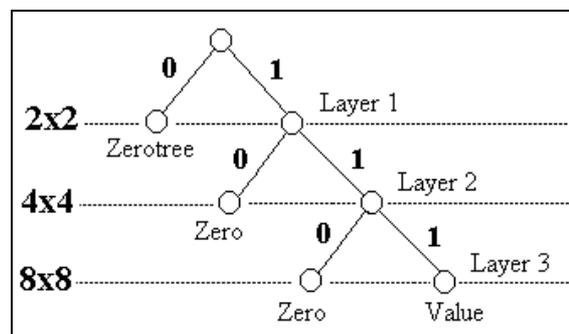


Figure 5. The HFSVQ block sizes and coding tree that are used by encoder.

There are some advantages for this zerotree subcodebook generation strategy;

- 1 Only layer members are involved in training for generation of the codebook. Both the iteration times and the number of comparison in each iteration are considerably reduced. As a result, the total training time is shortened.
- 2 By applying layer assignment, the size of each codebook is adjusted according to the accuracy requirement for reconstruction of the different regions of a video frame.

In the encoding process if the coefficients of a parent block in the coarsest subband are zero then this is defined as a zerotree and encoder passes to scan an other zerotree. If the coefficients of any subband block are zero then quadtree scan gives zero to this layer and changes the layer. If the blocks have no zero information then corresponding address is used instead of the original block. The more zerotrees and zerolayers the less bits encoder requires in the coding process.

The rate control is performed by the encoder in the three possible mechanism:

- 1 Increasing the number of the Lloyd-Max quantization intervals of any subband also increases the total bitrate and quality of the video frame. On the contrary decreasing it causes low quality and low bit rate (initial number of Lloyd-Max quantization interval is 10),

- Increasing the bandwidth of the deadzone area of any subband causes a decrease both in the quality of video frame and total bit rate. On the other hand smaller deadzone areas of any subband increases the bit rate and the quality of the frame. Note that increasing or decreasing the parameters mentioned above are performed in all subbands (initial area is 5 % of the maximum wavelet coefficient value),
- Increasing and decreasing the codebook sizes in each DWT level change the bit rate (initial codebook sizes for level 3=4, level 2=8, level 3=16).

### 3. SIMULATION RESULTS

In our first experiment, I frame coding results of ZTE, MPEG-4 and ZTW-HFSVQ were compared. The comparison was made at 14 kb, 27 kb and at 28 kb using the first frames of Akiyo, News, Foreman. The PSNR results for the luminance component (Y) and the average for the two chrominance components Cb and Cr labeled as C are shown in Table 1. The performance results of MPEG-4 and ZTE shown in Table 1 were given in (Sikora, 1997). It is seen that ZTE gives better performance than MPEG-4. On the other hand the proposed encoder seems to be better than the other two algorithms mentioned above both in luminance and chrominance for all test sequences.

Tablo 1. QCIF First I-Frame Coding PSNR Results in dB.

QCIF First I Frames	Bits (kb)	Y	MPEG 4	ZTE	ZTW HFSVQ
Akiyo	14	Y	33.06	34.62	34.62
		C	36.31	36.19	36.41
Akiyo	28	Y	38.42	40.18	40.82
		C	40.81	40.81	41.43
News	14	Y	28.60	29.38	29.42
		C	33.82	33.47	34.25
News	27	Y	33.38	34.49	35.13
		C	37.39	36.84	40.06
Foreman	14	Y	30.11	30.86	30.88
		C	38.27	38.69	39.42
Foreman	27	Y	35.05	35.27	36.16
		C	40.71	40.76	41.70

The second experimental result is a comparison of ZTE, MPEG-4 and ZTW-HFSVQ for P-frame coding using 100 frames of Akiyo QCIF video sequence at the rates of 5 fr/s and 10 kb/s. The simulation results for luminance and chrominance frames are shown in Figure 6 and Table 2. It is seen that, the ZTW-HFSVQ algorithm has 1.19 dB and 3.08 dB better performance than MPEG-4 and ZTE respectively for luminance frames and 2.34 dB and 3.96 dB for chrominance frames.

Table 2. P Frame PSNR Results For Akiyo at the Rates of 5 fr/s and 10 kb/s.

Sequence	Y	MPEG 4	ZTE	ZTW HFSVQ
Akiyo	Y	35.64	35.75	36.83
	C	39.48	39.86	41.82

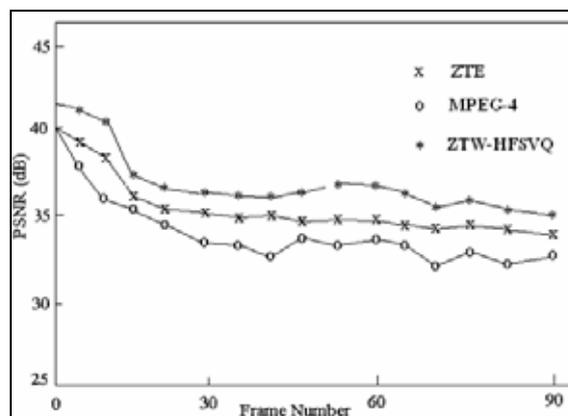


Figure 6a. Luminance PSNR Results For Akiyo at the Rates of 5 fr/s and 10 kb/s.

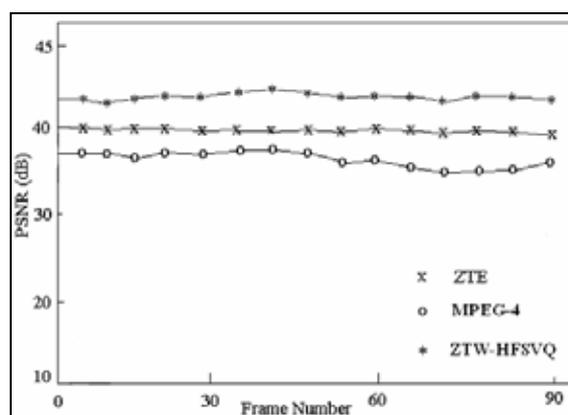


Figure 6b. Chrominance PSNR Results For Akiyo at the Rates of 5 fr/s and 10 kb/s.

The third comparison was made for 100 frames of Coastguard QCIF video at the rates of 7.5 fr/s and 48 kb/s. The comparison of ZTE, MPEG-4 and the proposed scheme, for encoding luminance and chrominance P frames is shown in Table 3. The average PSNR value of the ZTW-HFSVQ encoder over the entire coded sequence is 30.92 dB for luminance and 42.82 dB for chrominance frames. The average improvements in PSNR of the introduced encoder over MPEG-4 and ZTE are 1.18 dB and 1.72 dB respectively for luminance frames and 2.04 dB and 1.94 dB for chrominance frames. The interframe visual comparison of 40<sup>th</sup> luminance Coastguard frame between ZTW-HFSVQ and MPEG-4 is shown in Figure 7. When the PSNR

values are compared in Figure 7., it is seen that the ZTW-HFSVQ algorithm has 3.6 dB improvement over MPEG-4 algorithm.



(a)



(b)



(c)

Figure 7. Visual comparison of 40<sup>th</sup> luminance Coastguard frame a) original frame b) reconstructed by MPEG-4, PSNR=29.7 and c) reconstructed by ZTW-HFSVQ, PSNR=31.3 dB.

Table 3. P Frame PSNR Results For Coastguard at the Rates of 7.5 fr/s and 48 kb/s.

Sequence	Y C	MPEG 4	ZTE	ZTW- HFSVQ
Coastguard	Y C	29.74 40.78	29.20 40.88	30.92 42.82

It is seen from the simulations above that there are some performance differences between the video compression techniques. In slow motion videos such as Akiyo, News and Foreman, the luminance frames are better coded in ZTE technique than MPEG-4. In the fast motion videos such as Coastguard, on the contrary, the luminance frame coding performance of MPEG-4 is better than ZTE technique. On the other hand, the performance of the proposed new video compression technique in both slow and fast motion videos is better than MPEG-4 and ZTE.

#### 4. CONCLUSION

In this paper a video compression technique with comparable low bit rates is presented. The major components of the proposed encoder are the new diamond search block motion estimation, the overlapping block motion estimation, the discrete wavelet transform, the Lloyd-Max quantizer for quantizing the wavelet coefficients and the combination of hierarchical finite state vector quantization with the zerotree wavelet structure to encode the quantized wavelet coefficients.

The simulation results of the proposed video encoder was taken using the standard QCIF Akiyo, Coastguard, News and Foreman sequences. Average PSNR over the entire coded sequences shows that the proposed ZTW-HFSVQ video compression technique achieves comparable performance over MPEG-4 and ZTE both I and P frames.

#### 5. REFERENCES

- Auyeung, C., Kosmach, J., Orchard, M. and Kalafatis, T. 1992. "Overlapped Block Motion Compensation", SPIE'92, pp. 561-572.
- Kilic, İ., Yilmaz, R. 2003. "A Video Compression Technique Using Zerotree Wavelet and Hierarchical Finite State Vector Quantization", ISAPA 2003, pp: 311-316.

- Lin, Y., Tai, S. 1998. "A Fast Linde-Buzo-Gray Algorithm in Image Vector Quantization", IEEE Transactions on Circuits and System – II: Analog and Digital Signal Processing, Vol. 45, pp. 432-435.
- Lloyd, S. P. 1982. "Least Squares Quantization in PCM", IEEE Tran. On Information Theory, Vol. 28, pp. 129-137.
- Lu, J. and Liou, M. L. 1997. "A Simple and Efficient Search Algorithm for Block - Matching Motion Estimation", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, pp. 429-433.
- Martucci, S. A., Sodagar, I., Chiang, T. and Zang Y. 1997. "A Zerotree Wavelet Video Coder", IEEE Transactions on Circuits and System for Video Technology, Vol. 7, pp. 109 -118.
- Po, L. M. and Ma, W. C. 1996. "A Novel Four-Step Search Algorithm for Fast Block Motion Estimation", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 6, pp. 313-317.
- Rao, K. R., Hwang, J. J. 1996. "Techniques and Standards for Image, Video, and Audio Coding", Prentice Hall, New Jersey.
- Sikora, T. 1997. "The MPEG-4 Video Standard Verification Model", IEEE Transactions on Circuits and System for Video Technology Vol. 7, pp. 19-31.
- Yeung, E. 1997. "Image Compression Using Wavelets", CCECE 1997, pp. 241-244.
- Yu, P., Venetsanopoulos, A. N. 1994. "Hierarchical Finite State Vector Quantization for Image Coding", IEEE Transactions on Comm. Vol.42, pp. 3020 - 3026.
- Zhu S., Ma, K. K. 2000. "A New Diamond Search Algorithm for Fast Block Matching Motion Estimation", IEEE Transactions on Image Processing, Vol. 9, pp. 287-290.
- 
-