WZD – A Very Low Complexity Wavelet Video Codec for an ASIC Macro

W. C. Lynch, K. Kolarov, W. Arrighi, R. Hoover
Interval Research Corporation
1801Page Mill Road, Bldg. C, Palo Alto, CA 94304

Abstract

WZD (Wavelet Z-coDec) is a component or composite video codec engineered and optimized to greatly reduce the ASIC silicon implementation area in return for a modest degradation of the R-D (rate/distortion) trade-off. It consists of wavelet transforms organized into a carefully designed pyramid, dyadic quantization, and a novel application of Z-coding and Huffman coding to entropy code the wavelet coefficients. The resulting design can be completely implemented, including RAM, in less than 10% of the silicon area of and MPEG2 encoder excluding RAM.

1. Introduction

An image transform codec consists of three steps: 1) a reversible transform, often linear, of the pixels for the purpose of decorrelation, 2) quantization of the transform values, and 3) entropy coding of the quantized transform coefficients. This paper presents an entropy codec WZD which is fast, efficient in silicon area, coding-wise efficient, and practical when the transform is a wavelet pyramid [3]. We will focus on natural scene images quantized to match the human visual system (HVS) [7].

The primary features of WZD are:

- Enables extremely small and inexpensive ASIC macro implementations driven by the elimination of multiplies and divides and by the reduction of memory requirements.
- Reduced memory requirements by means of intermediate compression and entropy coding
- Simplified, quality block processing enabled by novel edge filters
- Integrated demodulation and processing of chroma directly from the composite signal
- Integrated, simple, inexpensive color rotation performed on the compressed signal
- Low power consumption due to a very low gate count and a very low (14.3 or 28.6 MHz) clock rate.

The WZD codec described applies to the (e.g., NTSC) composite video signal after demodulation from the RF carrier and after the FM modulated audio has been separated. The NTSC composite signal has YIQ chroma quadrature phase modulated onto a chroma subcarrier. WZD is capable of demodulating the chroma signals and of inexpensively rotating the chroma signals into other standard forms such as YUV or RGB.

2. ASIC Economics

The main drivers of silicon area (hence cost) in any implementation of a transform coder are the areas devoted to multiplies, divides, and to memory. WZD nearly eliminates multiplies and divides, executing the remaining ones as shifts. Memory utilization is reduced to the order of 1 megabit.

The economics of ASICs that pertain to our work are as follows. The size of the smallest feature in an ASIC is called λ. Typical values of λ are λ = 0.00035 mm (conservative), 0.00025 mm (contemporary), and 0.00018 mm (soon). A DRAM cell (1 bit) requires 30 λ^2. So 1.25 Mbits (see section 4) of DRAM requires 37.5 million λ^2.
A gate requires about 700 λ^2. So 40,000 gates (a reasonable estimate) requires another 28 million λ^2 for a total of about 65 million λ^2. At 0.00035 mm/λ, this is 8 mm^2. Logic silicon costs

about \$0.25/mm^2 so that 8 mm^2 costs \$2.00. The yield of good parts at this size is near 100% so losses from defective silicon should be small. The cost decreases as λ^2 so that at 0.00025 mm the cost is \$1.02 and at 0.00018 mm, \$0.53.

By way of comparison, the latest low cost MPEG2 encoder chips are ~100 mm^2 and require 4-8 Mbytes of external RAM. Using an optimistic Moore's law improvement rate of 40%/year it will take over 8 more years for MPEG2 encoders to achieve the estimated size of WZD.

3. Wavelets and Edge Filters

The basic wavelet used is the TS (2-6) transform, the quadratically lifted Haar wavelet.

$$
\begin{aligned}
f_i &= x_{2i} + x_{2i+1}\\
g_i &= x_{2i} - x_{2i+1}\\
h_i &= -\frac{f_{i-1}}{8} + g_i + \frac{f_{i+1}}{8}
\end{aligned}
$$

$$
\begin{aligned}
g_i &= \frac{f_{i-1}}{8} + h_i - \frac{f_{i+1}}{8}\\
x_{2i} &= \frac{f_i + g_i}{2}\\
x_{2i+1} &= \frac{f_i - g_i}{2}
\end{aligned}
$$

Figure 1. 2-6 forward and reverse transform.

This wavelet requires only multiplies and shifts for implementation. Similar wavelets have been used in the CREW work done at Ricoh of California [5]. The support of the wavelet is short which both reduces the computation and the extent of ringing from quantization errors in the wavelet coefficients. Furthermore, decode begins with a coarse-to-fine central quadratic interpolation so that quadratically smooth image areas are represented by the scale values f_i only (as seen in figure 1).

$$
\begin{aligned}
h_0 &= g_0 - \frac{3f_0}{8} + \frac{f_1}{2} - \frac{f_2}{8}\\
g_0 &= h_0 + \frac{3f_0}{8} - \frac{f_1}{2} + \frac{f_2}{8}
\end{aligned}
$$

Quadratic Filters

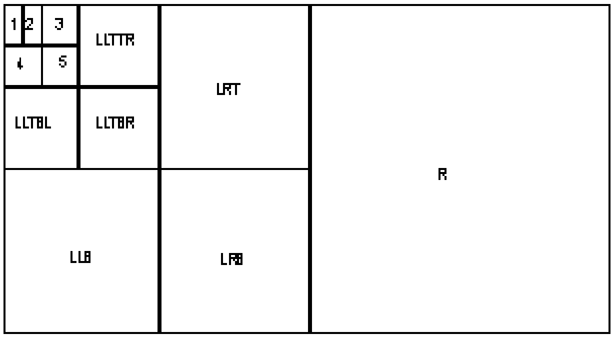
Figure 2. Left Edge Filters in WZD.

We have designed quadratic lifting formulas for the right and left (see figure 2) edges so that, if desired, block boundaries can be introduced while maintaining quadratically smooth interpolation across those block edges. As a result, blocking artifacts will appear only at extremely low bit rates. These edge filters can be implemented with adds, shifts, and a multiply by 3 (i.e., one add and one shift). In turn, block-by-block processing can substantially reduce DRAM utilization and reduce DRAM bandwidth requirements by simplifying memory access patterns.

4. The Component Video Pyramid

Incoming NTSC component video is digitally sampled (8 bits) at 14.3 MHz = 4f$_C$ where f$_C$ is the frequency of the chroma carrier [4]. Horizontal lines are cropped from 910 pixels to 704 (=22x32) pixels. As these 704 pixels stream in from the digitizer the first two levels of wavelet transform are applied in the horizontal direction, resulting in 11 bit values.

Within each of the two fields of a frame stripes of 8 lines each are double buffered into DRAM, requiring 123904 bits (2 stripes x 8 lines/stripe x 704 pixels/line x 11 bits/pixel) for the two stripe buffers. Each stripe is divided into 22 blocks, each block consisting of 8 lines and 32 columns. After the first two levels of horizontal wavelet (above) each block then consists of 8 lines, 8 columns, and 3 (horizontal) subbands. The first horizontal filter doubles the width of the scale pixels and compensates for the double height field lines.



1: LLTTLLTL (apex), 2: LLTTLLTR, 3: LLTTLRT,
4: LLTTLLB, 5: LLTTLRB

Figure 3. Component video Mallat pyramid(3x5) with 12 subbands

These blocks are read back from DRAM block-by-block, transformed internally by a 3-level wavelet pyramid, subjected to a preliminary (mild) quantization, entropy encoded (Z-coded), and written back to DRAM in linked buffers lists. The resulting pyramid for each of the component luma and chromas is depicted in figure 3, where R denotes the lifted difference and L the sum in the horizontal direction. B and T are the corresponding lifted difference and sum in the vertical direction (in figure 4).

Each luma block requires ~300 bits in its compressed form. Each field of 240x704 pixels = 30x22 blocks requires ~200000 bits of DRAM. Four fields of luma require ~800000 bits. The two chromas can be quantized so that they require only another 20% for a total field storage of ~1.0 Mbit. The luma stripe buffer adds 0.125 Mbit and the chroma stripes add another 0.125 Mbit. The total required RAM is thus 1.25 Mbit.

A group of pictures (GOP) consists of 4 fields or two frames. Two levels of wavelet (Haar and linearly lifted Haar) filters are applied in the temporal direction to improve the compression. Four entropy decoders decode four corresponding blocks from the four fields. The resulting four bit streams are kept logically synchronized, skipping over the substantial number of pixels where all four coefficients are zero. Where at least one is non-zero the wavelet transforms are processed with bit serial arithmetic at pixel rates (14.3 MHz). This results in four bit streams, one for each of the temporal subbands. Each bitstream is subjected to a final dyadic quantization (divide by a power of two – i.e., right shift). Each stream fills a tagged buffer and the tagged buffers are merged into the final compressed stream.

5. Composite Video Processing

Composite chroma processing is facilitated by WZD. The horizontal, vertical, and temporal wavelet filters, together with the $4f_C$ sampling, serve to separate the luma and the chroma carrrier and to demodulate two chroma signals into separate subbands. On the initial two levels of horizontal filtering, the luma is separated into the LL band while the two chroma signals appear in the LR and RR subbands, respectively. The RL subband contains mostly noise above 7 MHz. Subsequent vertical and temporal filters act as "comb" filtering, resulting in excellent chroma separation. The complete wavelet pyramid for composite video is shown in fig. 4.

Since the temporal filtering is performed bit-serial on transformed and compressed coefficients, there is the opportunity to perform the 3x3 matrix multiply for color rotation by means of three adders and a few registers configured to perform serial by parallel multiplies at the pixel rate. This is enabled by performing the rotation directly on the few non-zero wavelet coefficients.

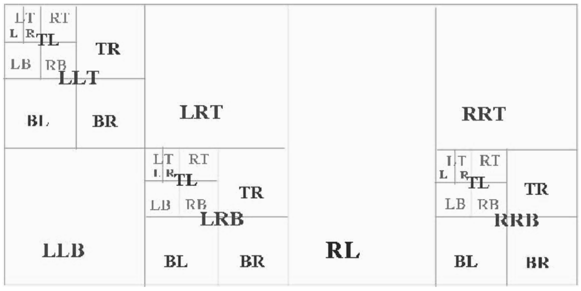


Figure 4. The composite video pyramid

6. Quantization and Entropy Coding

Dyadic quantization restricts the quantization divisors to powers of two. The result is that quantization can be performed by a right shift. Our experience and testing shows that this parameter space is more than adequate to achieve visually optimal values. Wavelets transform over a wider frequency range than is used in the DCT/JPEG scheme. This wider frequency range has a greater span in the contrast sensitivity curve for the human visual system (see [2, 7]), resulting in more design latitude in the quantization coefficients and less difficulty in selecting dyadic quantization coefficients.

The entropy codec involves a novel combination of the Z-coder and Huffman coding. This already has been reported. The Z-coder [1] is an approximate arithmetic coder for a binary code set that avoids multiplies and uses only adds, shifts, and lookups in modest size tables.

7. Results and Conclusions

Extensive tests of video picture quality have been conducted. The trade-off between MPEG and WZD is roughly the increased requirement for transmission and storage in return for extremely inexpensive encoding and decoding.

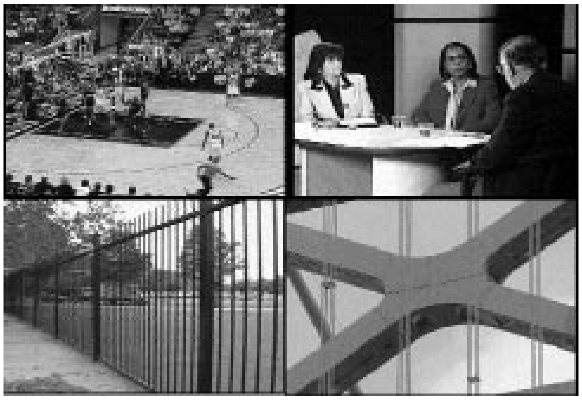


Figure 5. Video sequences used for testing

We tested the algorithm on several NTSC clips which vary in content and origin shown in Figure 5. The first one is a cable broadcast of an interview ("talking heads") without much motion. The second clip is a clean, high quality sequence from a laserdisk with a panning motion of a fence with vertical bars close together and motion of cars on the background. The next clip is a DSS (satellite) source - a basketball game (already MPEG2 compressed / decompressed) with a lot of motion and detailed crowd and field. The last clip is a high quality sequence from a laserdisk with a zooming motion on a bridge with a number of diagonal cables. The size of the frames is 720x486.

Figure 4 illustrates the signal-to-noise results for the four sequences. For comparison we have used a high quality commercially available MPEG2 codec from PixelTools. The MPEG2 was generated using the best possible settings for high-quality compression. We used 15 frames in a GOP (group of pictures), 3 B frames between I, P frames, 29.97 frame rate, 4:2:0 chroma format, medium search range double precision DCT prediction, stuffing enabled, motion estimation sub-sampling by one.

As we can see from table 1, our WZD coder is on par with MPEG.

	WZD 1.0 bpp	**MPEG2** 1.0 bpp	**WZD** 0.5 bpp	**MPEG2** 0.5 bpp
TalkShow	36.10 dB	37.03 dB	34.30 dB	34.85 dB
Fence	31.27 dB	29.62 dB	27.13 dB	26.00 dB
Basketball	28.33 dB	31.69 dB	25.33 dB	28.42 dB
Bridge	39.86 dB	39.35 dB	36.85 dB	37.34 dB

Table 1. WZD vs. MPEG2 commercial coder (from PixelTools) in signal-to-noise ratio (PSNR)

We were surprised by the general parity between MPEG2 and WZD in both PSNR and subjectively perceived picture quality, and particularly surprised by the clear superiority of WZD on the fence sequence

The WZD results in table 1 use an array of standard arithmetic coders [6] which are run separately on each bitplane of the wavelet coefficients.

8. Summary

We have described the low cost WZD codec, whose encoder is at least an order of magnitude less in silicon area and which achieves a rate-distortion result competitive with MPEG2. It has the additional ability to deal directly with composite video.

References

[1] L. Bottou, P. G. Howard, and Y. Bengio, The Z-Coder Adaptive Coder, *Proceedings of the Data Compression Conference, pp. 13-22,* Snowbird, Utah, March 1998.

[2] R. DeVore, B. Jawerth and V. Popov, Compression of wavelet decompositions, *Amer. J. Math*

[3] S. Mallat, A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, pp. 674-693,* 1989.

[4] C. Poynton, A Technical Introduction to Digital Video, *John Wiley & Sons, Inc.,* 1996.

[5] E. Schwartz, A. Zandi, and M. Boliek, Implementation of Compression with Reversible Embedded Wavelets, *Proceedings of the SPIE 40th Annual Meeting, Vol. 2564-04,* July 1995.

[6] I. Whitten, R. Neal, and J. Cleary, Arithmetic Coding for Data Compression. *Communications of the ACM 30, 6 (June) 1987, pp. 520-541*

[7] B. Wandell, "Foundations of Vision", Sinauer Associates, Inc. Publishers, 1995.