Analysis of Optimized Design Tradeoffs in Application of Wavelet Algorithms to Video Compression

Paul Wanis and John S. Fairbanks
Telemetry West
L-3 Communications Corporation
San Diego, CA 92123-3507

Cleared by DoD/DFOISR for public release under 04-S-2567 on 8/30/04

ABSTRACT

Because all video compression schemes introduce artifacts into the compressed video images, degradation occurs. These artifacts, generated by a wavelet-based compression scheme, will vary with the compression ratio and input imagery, but do show some consistent patterns across applications. There are a number of design trade-offs that can be made to mitigate the effect of these artifacts. By understanding the artifacts introduced by video compression and being able to anticipate the amount of image degradation, the video compression can be configured in a manner optimal to the application under consideration in telemetry.

KEYWORDS

Video Compression, Wavelet Transforms (WT), Flight Telemetry

INTRODUCTION

The optimization of video compression is a multi-function design problem with choices of different algorithms each producing varying image quality as a function of frame rate and signal-to-noise ratio, amongst others. Each of the design trade-offs represents characteristics about the compression which cannot be simultaneously maximized and thus must be sensibly traded-off. Some of the applications for compressed video, for which wavelet transforms are employed, are unmanned vehicles, teleconferencing, and digital library queries [1]. A central design dilemma in the use of wavelet algorithms for video compression is how to optimize the tradeoffs inherent in video compression for a customized application and limited bandwidth. Thus, a strategy for optimal design of compressed video by WT is presented with an aeronautical flight example of its implementation.

A WT is applied to video streams and uses estimation of changes in space and time of video streams to compress the data in the stream. Compression algorithms consist of four major steps: pixel shifting, data transform decorrelation, coefficient quantization, and coding. Within each part of the compression process, much research has been done and presented. A WT is applied to the data transform decorrelation of the video stream. The decorrelation of the compression process has spatial and temporal dependencies, which in turn affect the video quality upon decompression as a function of bit rate and signal-to-noise ratio.

WAVELET TRANSFORMS

A WT finds use in video compression because of their deterministic method over other compression algorithms. Commercially available products incorporate WT in video compression and have been described elsewhere in the literature [2]. Our goal here is to understand some of the implications of WT algorithms and current research for which optimal choices in the application and design of video compression systems can be made [3].

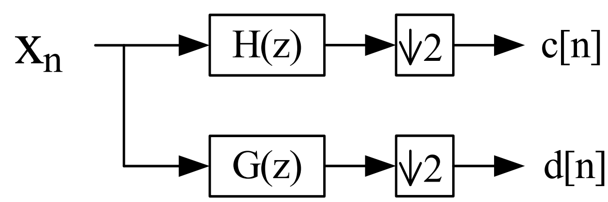

Fig. 1: Wavelet Transform[4]

A WT can be described in terms of filter banks which decompose data according to frequency bands that are spatially adjacent and globally optimized. The two branches in Fig. 1 represent the high, G(z), and low frequencies, H(z), of the image. The algorithm coefficients for scaling are c[n] and for the wavelet transforms are d[n] [4]. The frequency sub-band decomposition, in WT, which is spatially related, is shown in Fig. 2. This structured source and descendent relationship produces four spatially related descendents, however the number of levels of decomposition is variable.

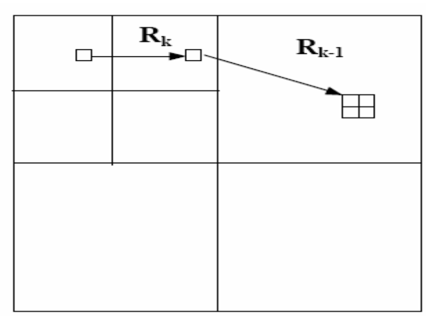

Fig. 2: Wavelet Transform

A key feature in WT as a multi-block decomposition in image processing is that the edges and unusual content can be processed in a similar manner to areas which have a large spatial continuity [5]. Thus, a scaling can be determined between space and frequency dependency in the image relative to significance for the decomposition components, as shown in Fig.3.

The general encoding scheme in WT involves three steps as shown in Fig. 4. The first step is a lossless transformation which produces symbols, each compared to the same threshold and statistically independent, with zero mean and symmetric distribution. The next step quantizes the symbols and relative to a threshold of bits per pixel (bpp).

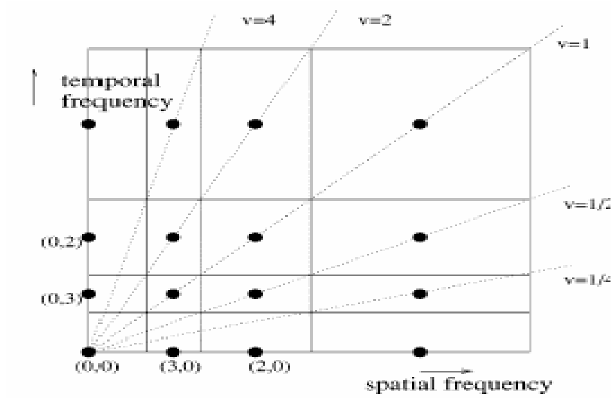

Fig. 3: Wavelet Decomposition Image Coefficient Relationship[6]

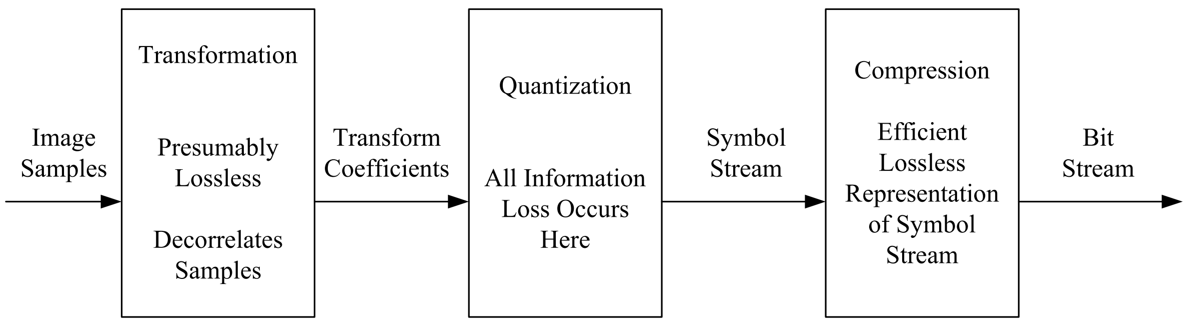

Fig. 4: Wavelet Encoding[5]

From this point an expression for the Entropy of the bit stream can be developed, as shown in (1), where p is the probability that a transfer coefficient became zero. The first

$$H = -p\log_2 p - (1-p)\log_2(1-p) - (1-p)[1-H_{nz}] \tag{1}$$

two terms in (1) can be taken as the Entropy of a significance map, while the last term is the conditional probability of non-zero bits. Significance mapping is the logging of the bit or symbol values. At this point, different compression algorithms are applied with different costs and efficiencies relative to a mapping of significance.

Finally the fidelity criterion or measure of the quality of a decompressed image is measured by the peak mean-square-root signal-to-noise ratio (PSNR), as shown in (2). Also, signal-to-noise ratio (SNR) and compression ratio (CR) are used to measure the performance of video compression [7]. The numerator is the maximum allowable signal amplitude, where B denotes the bit depth of the pixels. The denominator, which is also used to denote the noise power of the image, is the mean squared error between the processed image (signal plus noise) and the original image.

$$PSNR(dB) = 10\log_{10}\left(\frac{\left(2^B - 1\right)^2}{\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}(x,y) - f(x,y)\right]^2}\right) \tag{2}$$

Wavelet Transform Issues

Distortion in video compression occurs in the quantization portion of the process [8]. Each different strategy for improving the compression process produces different artifacts in the uncompressed picture.

Table 1: Order of Video Compression[8]

Wavlet decomposition $\rightarrow$ multiresolution motion compensation $\rightarrow$ multiscale quantization $\rightarrow$ entropy encoder or

Wavlet decomposition $\rightarrow$ multiresolution motion compensation $\rightarrow$ DCT $\rightarrow$ uniform quantization $\rightarrow$ entropy encoder or

Motion compensation $\rightarrow$ wavlet decomposition $\rightarrow$ multiscale quantization $\rightarrow$ entropy encoder or

Motion compensation $\rightarrow$ wavlet decomposition $\rightarrow$ DCT $\rightarrow$ uniform quantization $\rightarrow$ entropy encoder

From Table 1, the order of video compression can be seen to occur in different sequences. The effects of intraframe versus interframe compression which involve embedded zero tree wavelet (EZW) algorithm, set partitioning in hierarchical tree (SPIHT), and Flexible Block Wavelet Coding (FBWC) methods will be discussed [9]. The intraframe coding is referred to as the I-frame and the P frame in the predictive or interframe coding. In the case of EZW, a set threshold, θ, for the significance map is used but the descendents in the tree are also examined for significance in hierarchical manner. Thus EZW is a method for increasing the quality of the encoding according to a threshold where detail matters. In other words, a rule is used to pick up more detail for encoding when significant as opposed to treating all subdivisions equally.

In the case of SPIHT, which is similar to EZW, the difference is the magnitude ordering of the coefficients from the WT which improves the metrics of PSNR, bpp, and CR.

DESIGN IMPLICATIONS OF WT THEORY

Quality Boxes and Regions of Interest

In commercially available compression tools exist to specify a Region of Interest (ROI) function which allows direction of the partition schemes discussed above for emphasizing a path of partitioning relevant to the viewer. Another technique used, which is also commercially available, is termed the “quality box”. The quality box algorithm consists of attenuating the amplitude of the pixels outside the emphasized region prior to image compression, similar to the ordering algorithm discussed above in tree decomposition. The information content of the picture to be compressed is redefined for the WT; therefore, the efficiency of compression is improved. The sample images, left to right, of Fig. 5 show a compressed image without a quality box, followed next by image compression with a quality box applied with 18dB of attenuation of the imagery outside the emphasized area in the middle frame. In the third frame, image compression is shown using JPEG200 Region of Interest.

Fig. 5: Region of Interest Application

The PSNR is shown in Fig. 6, with the "normal" image unchanged between the interior and exterior of the emphasized area. The quality box approach results in a PSNR improvement between 3 and 5 dB within the emphasized area, but at the expense of a 20dB degradation in PSNR outside the image. Finally, the JPEG2000 ROI encoding shows a higher PSNR both inside and outside the emphasized region versus the quality box technique. Although JPEG2000 yields a higher PSNR outside the emphasized region than the quality box approach, observers may find the sharper resolution of the quality box approach more desirable.

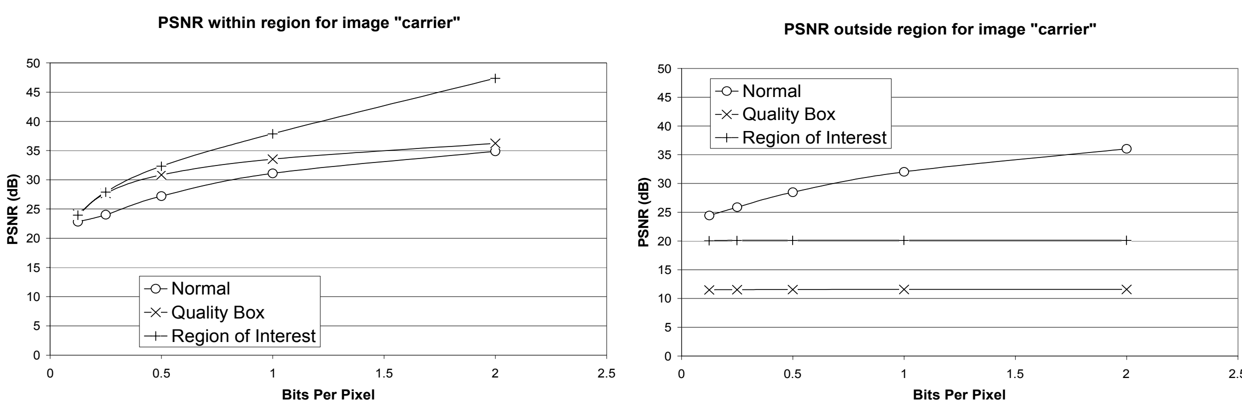

Fig. 6: PSNR of ROI Technique

Temporal Resolution

With intraframe WT compression, the bandwidth required to transmit the video imagery has a direct correlation to the video frame rate, shown in (3), where B is the required video data rate

$$B = H * V * D * F \tag{3}$$

(in bits per second), H is the horizontal resolution (pixels per line), V is the vertical resolution (lines per frame), D is the dynamic range of the frame (in bits per pixel) and F is the frame rate (in frames per second). Any reduction in the video frame rate will yield an identical reduction in the required transmission data rate. Alternatively, a reduction in the video frame rate can be used to gain more spatial resolution (in bits per pixel) at the same bandwidth.

Many different media employ different image update rates to produce visible motion. In all the media below, the imagery is perceived by the viewer to be moving despite the variety of update rates. Many applications, such as security cameras as well as older NASA footage, utilize lower frame rates and still provide all necessary temporal information to the viewer.

Table 2: Video Update Rates

Medium	**Image Update Rate**
Television (US)	30 Hz
Television (Europe)	25 Hz
Motion Pictures (US)	24 Hz
Animation (hand-drawn)	16 Hz

As a concrete example, the images shown in Fig. 7 were extracted from a movie “landing” with a temporal resolution of 30 frames per second. The four frames shown represent a total time of 133 milliseconds. The images clearly communicate the sequence of events.

Fig. 7: Full Rate Frame Sequence

In Fig. 8, the frames were decimated by a factor of 3 to 1, for a temporal resolution of 10 frames per second. The overall motion is still apparent in the video frames shown.

Fig. 8: 3 to 1 Rate Frame Decimation

Following further frame decimation the sequence below was decimated by a factor of 10 to 1, for a temporal resolution of 3 frames per second, as shown in Fig. 9. For this application, a 10 to 1 bandwidth gain can be realized with no appreciable information loss through frame decimation.

Fig. 9: 10 to 1 Rate Frame Decimation

Subsampling of Vertical Resolution

In standard NTSC and PAL video applications, each video frame consists of two interleaved video fields with one half the vertical resolution. Standard NTSC video has 486 visible lines per video frame at a rate of 30 frames per second. Each video frame is transmitted as two fields, the first field containing the odd lines the second field containing even lines. In some video compression applications, the video is compressed on a field-by-field basis at a rate of 60 fields per second. A simple way to improve bandwidth efficiency is to only use one field of video per video frame. The advantages of this technique are that it has no computational overhead, reduces required transmission bandwidth by a factor of two. A potential drawback of this direct technique is that it cuts the vertical resolution of the image in half.

Some sample images, shown in Figs. 10 and 11, show detail at full vertical and at one half vertical resolutions, respectively. In the image “Aldrin” the reduction in vertical resolution, while visible, does not impair comprehension of the image as there are no small items or diagonal lines in the field of view. In the image “carrier,” the reduction in resolution has a more noticeable impact. Smaller objects within this image show significant degradation, and the diagonal lines formed by the aircraft and its wings emphasize the reduced resolution.

Fig. 10: Excerpts from images “Aldrin” and “Carrier”

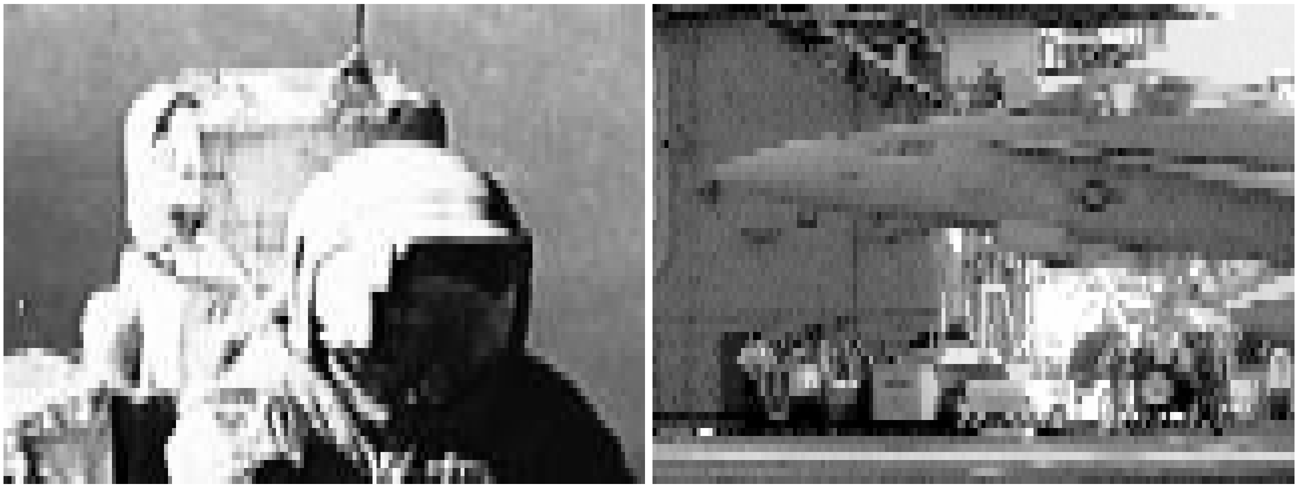

Fig. 11: Excerpts from “Aldrin” and “Carrier” after 2:1 vertical downsampling

The images were compressed using commercially available software and the PSNR of the resultant images is shown in Fig. 12.

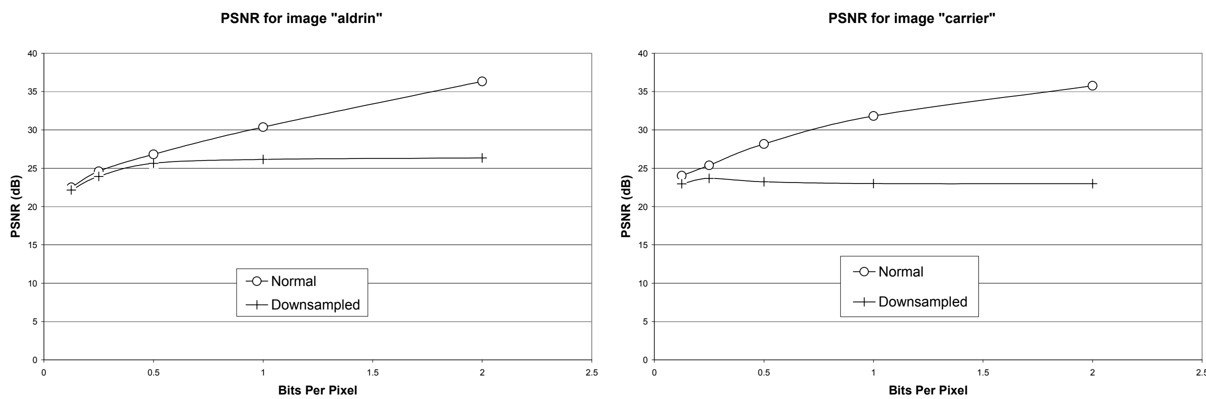

Fig. 12: PSNR for Aldrin and Carrier, Vertically Downsampled and Unmodified

The PSNR for the unmodified image degrades to roughly the same level as the down-sampled image when the compression level reaches 0.5 bits per pixel. At this point, we may infer that the quality of the imagery will not benefit from using two fields per frame of video, and so the bandwidth of the transmitted video data may be safely reduced by a factor of two without any noticeable impact on image quality.

In the image “carrier”, the PSNR for the down-sampled image is significantly lower than the full resolution image at most compression ratios. The two PSNR measurements do not converge until a compression ratio of 0.25 bits per pixel is chosen. It is clear that this application would not benefit from down-sampling, except at very high compression ratios.

Chrominance in Video

The use of EZW or SPIHT for I-frame encoding is based on gray-scale images [10]. Some color coding of video in performed as three separate gray-scale pictures with no benefit from the interdependent nature of the chrominance information. This is true of RGB or CMY systems. Some of the other schemes of representing color, involve luminance and chrominance (LC) combinations, such as YUV, which contains two chrominance and one luminance components [3, 5].

In many video monitoring applications the chrominance content of the imagery is unimportant to the observer and may be safely discarded. Monochrome imagery still maintains a high resolution and has a less information content to be compressed. The color decompositions of the image "bluang" into YUV components, using a forward irreversible component transform available in commercial software, are shown in Fig. 13. It is apparent that the chrominance (U and V) channels of the image have a lower dynamic range; therefore, less information content exists in the luminance (Y) channel. Table 3 lists the entropy of each channel and the total entropy, which accounts for the U and V channels having one-half the horizontal resolution of the Y channel as in standard CCIR-601 or CCIR-656 video applications. The entropy measurements show that the luminance information represents roughly 62% of the sample image information in Fig. 13. Discarding the chrominance information will reduce the information content for compression to 62% of its former content. As Fig. 14 indicates, an equivalent PSNR can be obtained with one half the bits per pixel, which would also have the benefit of cutting the video data rate in half.

Fig. 13: bluang Chrominance channels.

Table 3: Chrominance Content

Channel	Entropy (bits / sample)	Samples / pixel	Bits / pixel
Y	4.9	1	4.9
U	2.9	0.5	1.45
V	3.1	0.5	1.55
Total			7.9

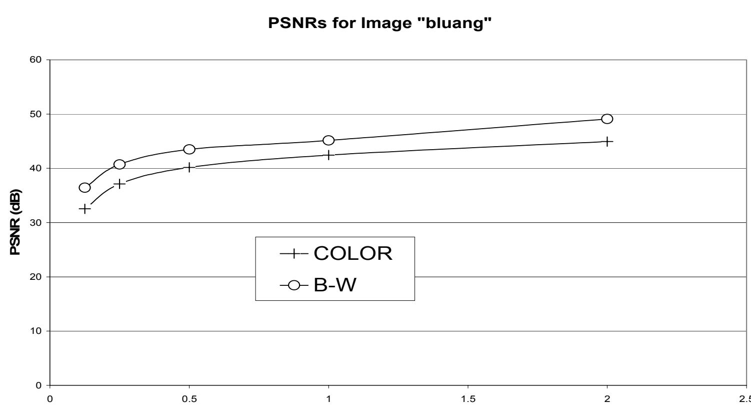

Fig. 14: Chrominance PSNR vs. bpp.

CONCLUSION, FUTURE RESEARCH

All video compression algorithms introduce artifacts from the uncompressed video as a function of the method employed. Several examples have been reviewed in this paper involving ROI, quality boxes, chrominance, temporal, and vertical resolution. Research in wavelet algorithm application to video compression provides for control of the compression process both by spatial and frequency distribution throughout a picture. Different processes decompose, resolve, quantize, and encode with differing costs in CR, PSNR, SNR, and subjective appearance. Similarly, exploiting correlations between video frames in both content and color can lead to significant reduction in necessary data transmission with limited channel capacity, as is the situation for telemetry. Future research on improvements in the compression algorithms can provide additional control parameters built into COTS for telemetry applications. With an increased awareness of the flexibility and its limits in current commercial products, better design choices in video compression can be made.

REFERENCES

[1] **S. Kandadai, "Object Detection and Localization in the Wavelet Domain," presented at International Telemetering, San Diego, CA, 2002.**

[2] **C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview,"** ***Consumer Electronics, IEEE Transactions on*****, vol. 46, pp. 1103-1127, 2000.**

[3] **R. C. Gonzalez and R. E. Woods,** ***Digital Image Processing*****, 2nd ed: Addison-Wesley, 2002.**

[4] **R. L. Claypoole, G. M. Davis, W. Sweldens, and R. G. Baraniuk, "Nonlinear wavelet transforms for image coding via lifting,"** ***Image Processing, IEEE Transactions on*****, vol. 12, pp. 1449-1459, 2003.**

[5] **J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients,"** ***Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*****, vol. 41, pp. 3445-3462, 1993.**

[6] **I. K. Levy and R. Wilson, "Three-dimensional wavelet transform video coding using symmetric codebook vector quantization,"** ***Image Processing, IEEE Transactions on*****, vol. 10, pp. 470-475, 2001.**

[7] **Q. Du and C.-I. Chang, "Linear Mixture Analysis-Based Compression for Hyperspectal Image Analysis,"** ***Geoscience and Remote Sensing, IEEE Transactions on*****, vol. 42, pp. 875-891, 2004.**

[8] **Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression,"** ***Circuits and Systems for Video Technology, IEEE Transactions on*****, vol. 2, pp. 285-296, 1992.**

[9] **B. S. Nanda and N. Kaulgud, "Effect of quantization on video compression," presented at Industrial Technology, 2002. IEEE ICIT '02. 2002 IEEE International Conference on, 2002.**

[10] **K. Shen and E. J. Delp, "Wavelet based rate scalable video compression,"** ***Circuits and Systems for Video Technology, IEEE Transactions on*****, vol. 9, pp. 109-122, 1999.**