IMPROVED ENHANCEMENT LAYER VIDEO CODING

By

RATHILALL SEWSUNKER

A thesis submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

WASHINGTON STATE UNIVERSITY

School of Electrical Engineering and Computer Science

August 2000

To the Faculty of Washington State University:

The members of the Committee appointed to examine the thesis of Rathilall Sewsunker find it satisfactory and recommend that it be accepted.

_____
Chair

_____

_____

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor T.R. Fischer, for his guidance and assistance during my study. He is an excellent teacher and mentor, and has patiently guided me through the graduate program. Sincere appreciation also goes to Dr B. Belzer, Dr K. Sivakumar and Professor C. Hsu, both for their well-taught classes as well as their assistance in serving on my committee. Thanks to Ruby Young for all her wonderful help.

I would like to thank the School of Electrical Engineering and Computer Science for the education that I have received. Thanks to my sponsor, the Africa-America Institute for providing me with a full scholarship and for their caring attitude during my study program. Particular thanks for that goes to Marks Chabedi, my Program Officer. Thanks to my employer, the University of Durban-Westville, for allowing me study leave, and especially to the Head of Electrical Engineering, Professor S. H. Mneney for his help and support.

Brian Banister has helped tremendously practically from my first semester at Washington State University through until the end. Brian's selfless disposition of helping and teaching will always be admired. Thanks to Esteban Rodriguez-Marek, and Ping Hou for their enthusiastic help with course-work and research alike. Thanks to my office-mate, Runar Solli, for his help and his jovial company.

Thanks to my Mum and my family in South Africa for their encouragement and support,

and especially to Kennybhai and Jaybhai, who took care of all kinds of responsibilities, not excluding sending money whenever I needed it!

IMPROVED ENHANCEMENT LAYER VIDEO CODING

Abstract

by Rathilall Sewsunker, M.S.
Washington State University
August 2000

Chair: Thomas R. Fischer

Scalability techniques are increasingly applied to video coding to alleviate transmission problems such as restricted bandwidth, network congestion and data loss. Being layered, scalable video offers the receiver a choice of bit-streams based on resources and needs. The thesis focuses on SNR scalability. The conventional method for SNR scalability provides a base layer description using coarse quantization, then encodes the refinement information using a smaller step size. For the case of a two-level encoder, the refinement information is simply a difference between the original data and the base layer description. This work illustrates that picture quality can be improved by encoding the enhancement information based on the value returned by the base layer quantizer, rather than as a blind difference. The idea is that for the two cases of the base layer quantizer returning a zero or a nonzero index, the probability density functions (pdfs) of the quantization error are different. Encoding the enhancement information using this method significantly improves the enhanced picture SNR, while also reducing the encoder complexity. This technique can also be applied to spatial scalability.

Contents

List of Figures

Dedication

This thesis is dedicated to my Spiritual Master,

Srila Bhakti Caru Swami Maharaja,

who encouraged me from the very first day to embark on this adventure,

and

to my wife, Ashika, and daughter, Krishnarupa, who have sacrificed so much just to be with me during this study.

Chapter 1

Introduction

1.1 Overview of Video Compression

The need for compressing video data efficiently is an important aspect of today's information-oriented living. The trend is that since more information is conveyed in a shorter time using the video medium, it is becoming increasingly popular. In terms of extensive biographical information, for example, many people would prefer watching a video, than reading a book. Whether it is viewing a video sequence on the Web, or just transmitting video data from one point to the next, compression lends a great helping hand in a world of limited and shared resources. Video is perhaps the source of the largest amounts of data and the goal of video compression, often called video coding, is to keep only that data which is necessary to replicate the video at the receiver at a quality acceptable to the end user. In view of the varying resource constraints such as available bandwidth and network congestion as well as varying user requirements, there is a growing need for scalable video coding. This means that several layers of video are simultaneously encoded, so that the receiver can choose the

decoded video quality, based on available resources or needs. In the following paragraphs the fundamental aspects of video coding are discussed. More details on scalable video are provided in Chapter 2.

In image compression the key strategy is to remove spatial redundancy. Video is essentially a sequence of images, and video compression attempts to remove both spatial and temporal redundancies. In many video sequences, while there may be some movement, a significant part of the image remains the same from frame to frame. In order exploit this redundancy, most video coding algorithms subtract a prediction off the current frame before encoding it. This prediction is computed from the previous frame, and it compensates for motion between the frames. This means that if there was only translational motion from one frame to the next, then no additional data need be sent to the receiver, except the motion vector data. The decoder then reconstructs the current frame using the motion vector plus the previously decoded frame.

Figure 1.1 shows the main operations in a video encoder. Motion-compensated prediction reduces the temporal redundancy. Transform coding, such as the Discrete Cosine Transform (DCT) reduces spatial redundancy by compacting the pixel energy into a few transform coefficients. Quantization represents the transformed coefficient values as indices, based on a step-size parameter, thereby further compressing the data. Entropy coding utilizes the statistics of the quantization indices to efficiently encode them. Entropy coding is lossless. Quantization, however introduces some distortion, and the process of video compression is essentially a trade-off between transmission rate and distortion at the

receiver. At the decoder the process is reversed.

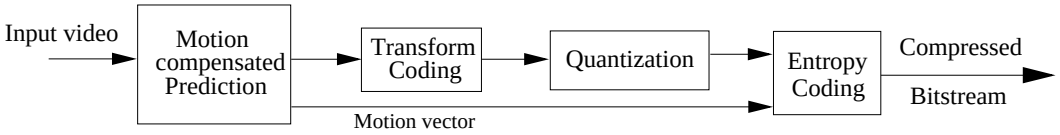

Figure 1.1: Block diagram of a video encoder.

1.2 Overview of the Thesis

The following four chapters form the thesis work. Chapter 2 gives a more detailed explanation of scalable video coding. It includes an overview of the scalability mode in the H.263+ video coding standard. Specifically, the SNR scalability mode is discussed in detail. Chapter 3 introduces the thesis contribution, namely the improved enhancement layer video coding method. It compares the new enhancement coding method to the existing method used in H.263+. A section discusses the use of trellis coded quantization (TCQ) [11] instead of scalar quantization (SQ) for the refinement coding, and presents some results of applying the new SQ-TCQ scalable coding method to memoryless data. Chapter 4 discusses the results obtained from applying the new method to the SNR scalability mode in H.263+, for SQ-SQ encoding. The results are compared to that of the existing H.263+ enhancement method. Finally, Chapter 5 presents the conclusions of the research and discusses areas of future work.

Chapter 2

Scalability Mode in H.263+

2.1 Introduction

H.263 Version 2, popularly known as H.263+ is a standard for video coding at low bit rates. The standard specifies a coding method for the compression of the moving picture component of audio-visual services. Applications include video conferencing and video telephony. H.263+ offers several optional modes that broaden the range of application of the standard as well improve its compression capability. An important one, especially in the context of congested, packet-lossy and heterogenous networks, is the scalability mode. The scalability mode may also be used in conjunction with error control techniques to improve codec performance. The next section discusses the topic of scalable video. Section 3 outlines the three scalability methods supported by H.263+. Section 4 gives a more detailed explanation of SNR scalability, after introducing some details of the baseline H.263+ codec.

2.2 Scalable Video

In scalable video coding, the first layer (base layer) is coded independently and additional layers (enhancement layers) are coded with respect to the base layer. The enhancement layers, therefore, reuse the bandwidth assigned to previous layers. There are three main scalability types; SNR scalability, spatial scalability, and temporal scalability. SNR scalability is discussed in more detail. A brief discussion of the other types follows.

2.2.1 SNR scalability

For simplicity, the discussion is restricted to one enhancement layer. An SNR scalable encoder provides two layers of video at the same spatial resolution, but at different qualities or SNR values. Figure 2.1 shows the conventional method used for SNR scalability. The base layer is the result of encoding at some coarse step-size, say Δ. The difference between the base layer reconstruction and the original data is encoded using a smaller step-size, say $\Delta/2$. The bit-streams for the layers are multiplexed and transmitted to the decoder. At the decoder the base layer is decoded, then enhanced by adding the refinement from decoding the second layer. Typically, there is a loss in SNR in using scalable coding compared to non-scalable coding at the same overall rate [14]. The important advantage of the use of SNR scalability, however, is the flexibility of transmission, and the ability to recover from transmission errors. For example, network bandwidth fluctuations may preclude the transmission of a fine non-scalable description. A scalable description can then make optimal

use of the resources. In the case of channel errors or packet losses, the base layer can be decoded and displayed by itself.

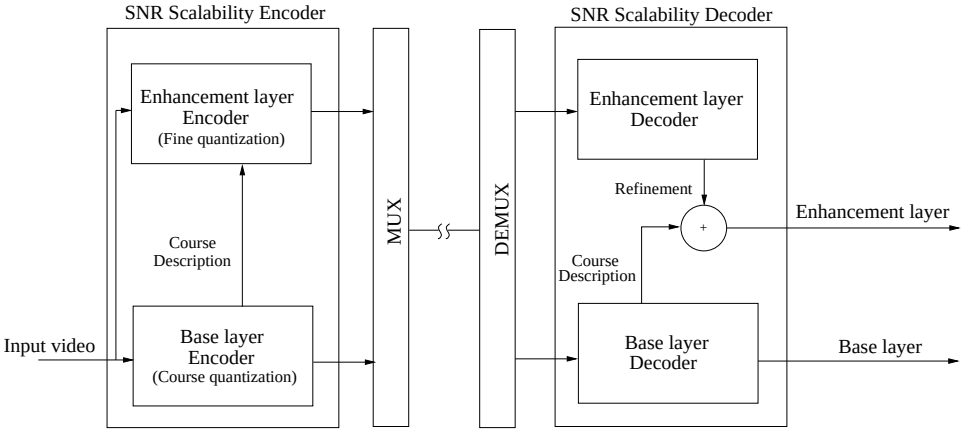

Figure 2.1: Block diagram of an SNR scalability codec.

2.2.2 Spatial scalability

In spatial scalability encoding, the base layer uses a lower spatial resolution than the enhancement layer. The enhancement layer is coded using a spatially interpolated base layer. Error resilience is achieved in spatial scalability coding by sending the critical lower layer data over a channel with better error performance.

2.2.3 Temporal scalability

Temporal scalability coding produces a base layer and a temporally higher resolution next layer. When combined the layers can provide the full temporal resolution available in the input video.

2.3 SNR, Spatial and Temporal Scalability Mode in H.263+

In H.263+ scalability mode, a base layer description is sent followed by one or more enhancement layers. This allows the decoding of a sequence at more than one quality level. In addition to I (Intra) pictures and P (Prediction) pictures, scalability uses three additional picture types, namely, B (bi-directional), EI (Enhanced Intra) and EP (Enhanced Prediction) pictures.

2.3.1 SNR scalability

Lossy compression introduces artifacts and distortion, and the difference between a reconstructed picture and its original, called the coding error, is almost always a non-zero valued quantity. Normally this coding error is lost at the encoder and never recovered. SNR scalability coding attempts to recover some of this loss by subsequently coding the error. This arrives at the decoder as an enhancement to the base layer description. The enhancement serves to increase the SNR of the reconstructed video, hence the name SNR scalability. Figure 2.2 shows the flow of information for SNR scalability. In going from a picture in the base layer to a picture in the enhancement layer, a finer quantizer is used to encode the error information. No motion vector information is needed. An enhancement layer picture may also be predicted from a previous enhancement layer picture. In this case motion information is necessary. Finally an enhancement picture may be bi-directionally predicted from both a prior enhancement layer picture and a temporally simultaneous base layer pic-

ture. In the implementation, the choice of which prediction is used is based on the lowest sum-of-absolute-differences (SAD) number.

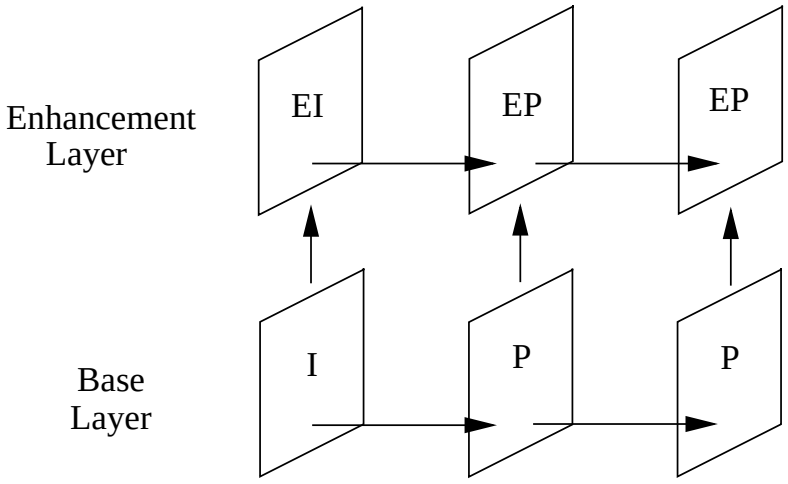

Figure 2.2: Data flow in H.263+ SNR scalability.

2.3.2 Spatial scalability

Spatial scalability is similar to SNR scalability. The difference is that before the picture in the base layer is used to predict the picture in the enhancement layer, it is interpolated by a factor of two either horizontally, vertically or in both directions. For example, a QCIF format (176×144) base layer picture would give a CIF format (352×288) enhancement layer picture using 2-D spatial enhancement. Except for requiring an upsampling process to increase the size of the reference layer picture prior to its use as a reference for the encoding of the enhancement, the processing and syntax for a spatial scalability picture is functionally identical to that for an SNR scalability picture. Figure 2.3 shows the flow of information for spatial scalability.

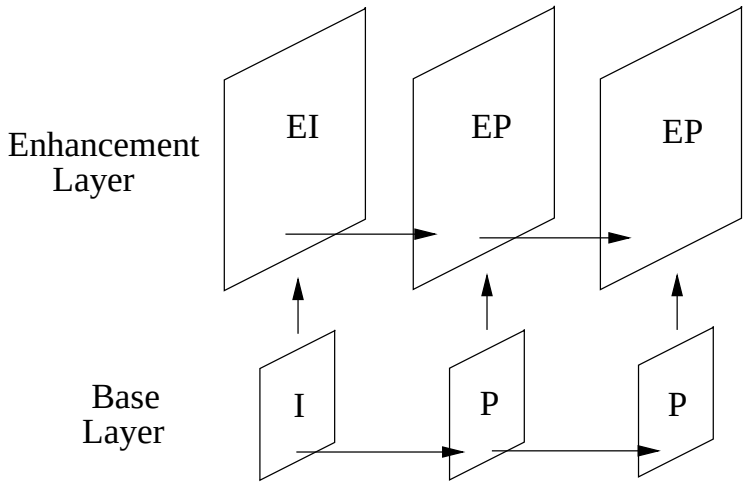

Figure 2.3: Data flow in H.263+ spatial scalability.

2.3.3 Temporal scalability

Temporal scalability uses bi-directionally predicted, or B pictures allowing the prediction from either one or both of a previous and a subsequent reconstructed picture in the reference layer. This results in better compression performance compared to the use of just P pictures [7]. B pictures are not used as reference pictures for the prediction of other pictures. This allows B pictures to be discarded if necessary without affecting subsequent pictures, thereby providing temporal scalability. Figure 2.4 shows the predictive structure of B pictures. If the pictures of the original video sequence were numbered $1, 2, 3, \ldots$, then the bitstream order of the encoded pictures would be $I_1, P_3, B_2, P_5, B_4, \ldots$. In the reference or base layer only the I and P frames are encoded. The enhancement layer codes the B frames. Any number of B pictures can be inserted between reference pictures, up to the full temporal resolution of the input video.

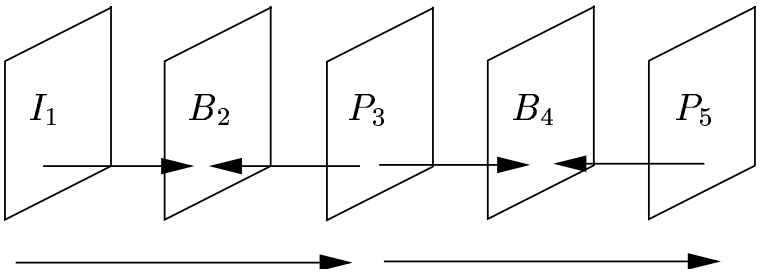

Figure 2.4: B picture prediction in H.263+ temporal scalability.

2.4 Detailed Discussion of SNR Scalability

In order to describe in detail the processes involved in SNR scalability coding, it is necessary to overview the baseline H.263+ codec.

2.4.1 Baseline H.263+ Codec

Figure 2.5 shows a simplified block diagram of the baseline H.263+ encoder. The baseline encoder uses motion-compensated prediction to reduce temporal redundancy and the Discrete Cosine Transform (DCT) to reduce spatial redundancy. The DCT coefficients are scalar quantized. The first frame is encoded without prediction, called an I picture. Subsequent frames are encoded as Inter or P pictures unless there is a drastic scene change in the input sequence, in which case another I picture is coded. The entropy code is a Variable Length Code (VLC) designed to match the statistics of the quantizer indices.

The encoder processes the input video in blocks called Macroblocks(MBs). Using an input video sequence in QCIF format (176 pixels by 144 lines), the MBs are organized as shown in Figure 2.6.

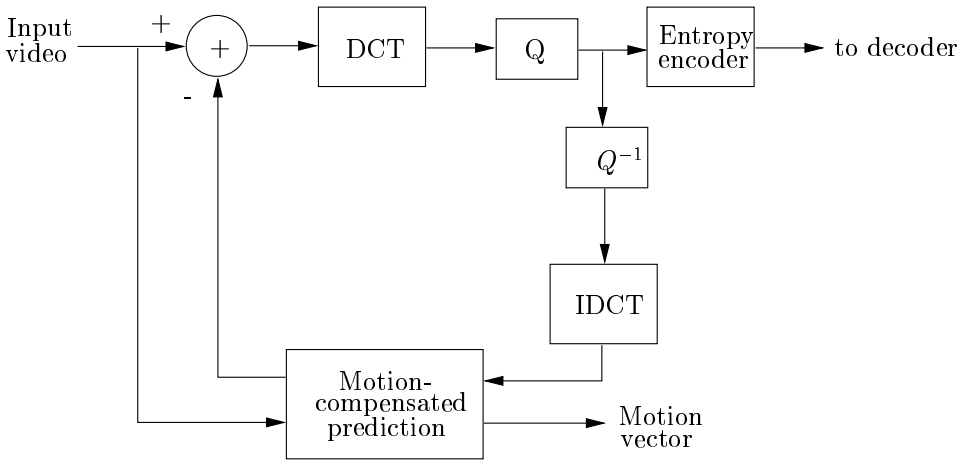

Figure 2.5: Simplified Block Diagram of the H.263+ Baseline Encoder.

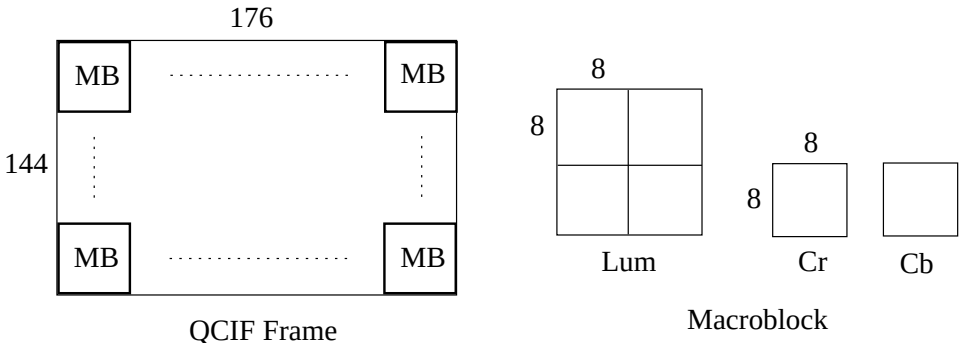

Figure 2.6: Macroblock arrangement for QCIF format.

In Inter mode, the prediction error frame, that is, the difference between the original frame and the motion compensated prediction frame, is encoded. The motion estimation searches out a fixed translational motion between the current and the previous picture. The motion information is transmitted to the receiver as a 2-dimensional motion vector (MV). The most widely used cost function for block matching algorithms is the sum-of-absolute-

differences (SAD) given by

$$SAD = \sum_{k=1}^{16} \sum_{l=1}^{16} |B_{i,j}(k,l) - C_{i-u,j-v}(k,l)| \tag{2.1}$$

where $B_{i,j}(k,l)$ represents the $(k,l)^{th}$ pixel of a MB from the current picture at the spatial location (i,j) and $C_{i-u,j-v}(k,l)$ represents the $(k,l)^{th}$ pixel of a candidate MB from the previous picture at the spatial location (i,j) displaced by the vector (u,v). The standard uses interpolation to compute half pixel prediction. The MB producing the smallest SAD within a search window is selected and the MV generated.

The H.263+ encoder uses several quantization rules. The Intra DC coefficients are quantized using a fixed step size of 8, according to,

$$LEVEL = (COF + 4)/8 \tag{2.2}$$

Intra non-DC and Inter coefficients are quantized using an even step size between 2 and 62, according to the following rules,

Intra non-DC:

$$|LEVEL| = |COF|/\Delta \tag{2.3}$$

Inter:

$$|LEVEL| = (|COF| - \Delta/4)/\Delta \tag{2.4}$$

where integer arithmetic is assumed.

Deadzone refers to the increase in the quantization bin around zero compared to other bins. Note that Inter coefficients are quantized using an increased deadzone, compared to Intra coefficients. The result of increased deadzone is that more coefficients are quantized to zero. This decreases the rate due to longer runs of zero-valued indices, but may also reduce picture quality.

In order to improve efficiency in H.263+, the quantization operation is implemented as a table look-up. The quantized coefficients are zig-zag positioned in an 8×8 block [7] and then entropy coded. The entropy coding is a combination of Huffman coding and run-length coding. Each block has a few nonzero levels separated by runs of zeros. The most common run-level combinations are assigned codewords in a frequency table. This table is empirically created based on the statistics of the quantizer indices. Each nonzero level together with its preceding run of zeros is then encoded using the frequency table. If the run-level is not a frequently occurring one, (i.e. not found in the frequency tables), 22 bits are used to represent it. A flag marks the last nonzero level. This is encoded using a different frequency table. Motion vectors and additional header information are added and the bitstream is transmitted to the receiver.

In the local loop, inverse quantization and inverse DCT (IDCT) are used to generate a reconstructed picture. For the inverse quantization, if the LEVEL = 0, the reconstruction value REC = 0. The reconstruction values for non-zero indices are given by,

Intra DC:

$$REC = LEVEL \times 8 \tag{2.5}$$

All other coefficients:

If QUANT = odd

$$|REC| = QUANT \times (2 \times |LEVEL| + 1) \tag{2.6}$$

If QUANT = even

$$|REC| = QUANT \times (2 \times |LEVEL| + 1) - 1 \tag{2.7}$$

where QUANT = $\Delta/2$. The above rule disallows even-valued reconstruction values. This has been found to prevent accumulation of IDCT mismatch errors [7]. After calculating |REC|, the sign information is added according to

$$REC = sign(LEVEL) \times |REC| \tag{2.8}$$

The reconstruction levels of all coefficients other than the IntraDC are clipped to the range -2048 to +2047. The 8×8 blocks are then passed through a two-dimensional IDCT. For the case of an Intra frame, this gives the reconstructed picture. For Inter frames, the motion-compensated prediction is added to give the final predicted picture for the next frame.

2.4.2 SNR scalability in H.263+

In this subsection the existing method of implementing SNR scalability in H.263+ is discussed. Figure 2.7 shows the block diagram of the existing method. For simplicity, the figure is based upon coding the I and EI frames only. Furthermore, it is assumed that all other optional modes are turned off. For the purpose of the investigation, the base layer quantizer step-size was set to Δ and the enhancement layer step-size set to $\Delta_{refine} = \Delta/2$. In order to obtain the coding error, the standard computes a pixel-domain difference between the original picture and the reconstructed base layer picture. The DCT is taken and the error is then encoded at the smaller step-size. Both the base layer and the enhancement layer step-size values are transmitted to the decoder as side information.

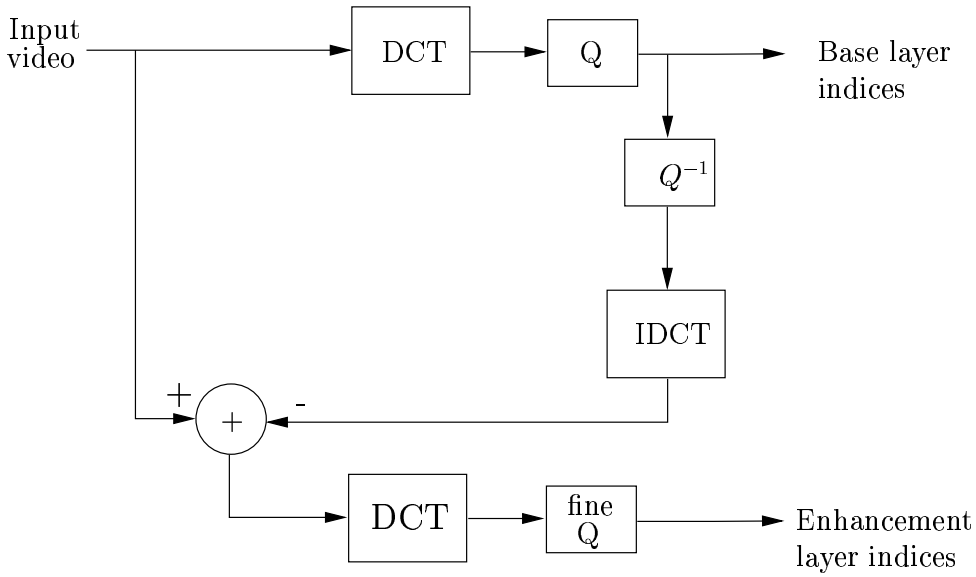

Figure 2.7: Block diagram of the existing method used for SNR scalability.

For enhancement quantization, all the DCT coefficients are treated as Inter coefficients,

and hence quantized with an increased deadzone as discussed in the last section. It is significant to note that for the case of the typical refinement step-size chosen, any non-zero index in the base layer, results in a zero index in the enhancement layer. This is one of the factors that makes the existing method sub-optimal. Only those zero-valued indices in the base layer with sufficiently large DCT coefficients in the enhancement layer enjoy refinement.

P frame enhancement is a little more complicated. In the base layer the motion-compensated prediction is subtracted off the current frame before encoding it. In the enhancement, one of upward, forward or bi-directional prediction is used, based on the lowest SAD number calculated. Upward prediction is favored by subtracting 50 from its SAD. The forward prediction SAD is unchanged, and the bi-directional SAD is penalized by adding 100. For the case of upward prediction, the pixel domain difference between the original picture and the base layer reconstruction is encoded using a step-size of Δ_{refine}.

Chapter 3

The Improved Method and Application to H.263+

3.1 Introduction

In this chapter the key thesis contribution, namely the new method for enhancement coding of video sequences, is presented. Section 2 gives the details of the new method and section 3 shows how the new method is applied to SNR enhancement in H.263+. Section 4 presents the case of using TCQ instead of SQ for the enhancement quantization. The results of applying the new SQ-TCQ codec to memoryless data are included.

3.2 Details of the New Method

The new method for enhancement coding can be applied to both SNR and spatial scalability types. The discussion is focussed on SNR scalability, since this is the mode that was investigated. The key concept in coding the enhancement in SNR scalability is that of differential coding, coding the error between the base layer description and the original

data. Figure 3.1 shows a two-stage differential encoder.

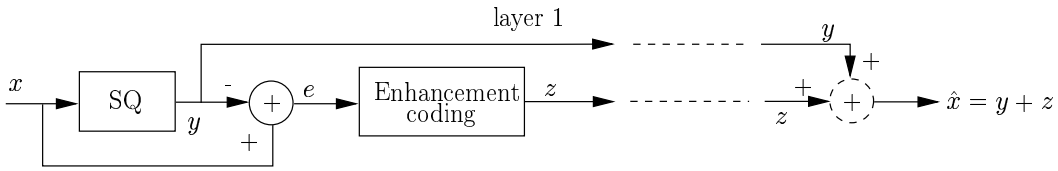

Figure 3.1: Two-stage differential encoder.

A simple example illustrates the shortcoming of the differential encoding method in the figure. Consider a source x with peaked density, for example Laplacian, as shown in Figure 3.2. It is desired to quantize the data using a uniform scalar quantizer with step-size Δ and deadzone as shown. Let y denote the midpoint reconstruction values.

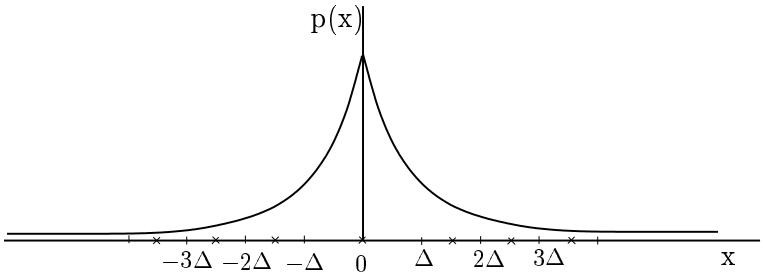

Figure 3.2: Laplacian source density.

The quantization error,

$$e = x - y \tag{3.1}$$

has a probability density function given by the composite of three cases as shown in Figure 3.3, based on y.

Clearly, a better encoding of the error can be achieved by conditioning on the value of y for the three cases $(y < 0, y > 0, y = 0)$. This information is available as the base

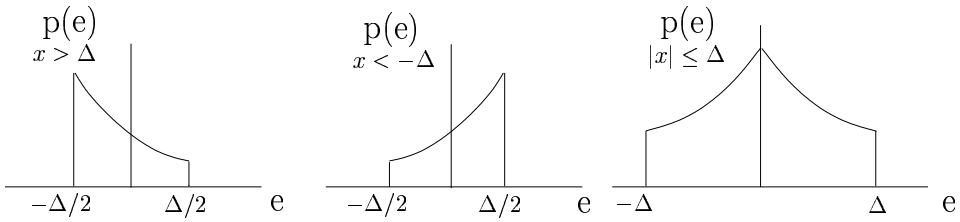

Figure 3.3: Three cases of error pdf based on y.

layer quantization indices, both at the encoder and the decoder. The error is conditionally encoded as follows:

For $y \neq 0$, encode

$$e = |x - q_{min}(y)| \tag{3.2}$$

where $q_{min}(y)$ is the lower (in magnitude) boundary of the quantization interval corresponding to the value of y.

For $y = 0$, encode

$$e = x - y = x \tag{3.3}$$

Although the sign information of the error is lost in equation 3.2, it can be recovered at the decoder as follows.

$$\hat{x} = q_{min}(y) + sign(y) * z \tag{3.4}$$

where z is the enhancement reconstruction.

3.3 Application to H.263+

Figure 3.4 shows the block diagram of the new method for SNR scalability applied to H.263. For simplicity, I frame encoding is shown. In the figure, Δ is the base layer quantization step-size. Uniform scalar quantization is applied to both the base layer as well as the refinement coding with deadzone as in H.263+.

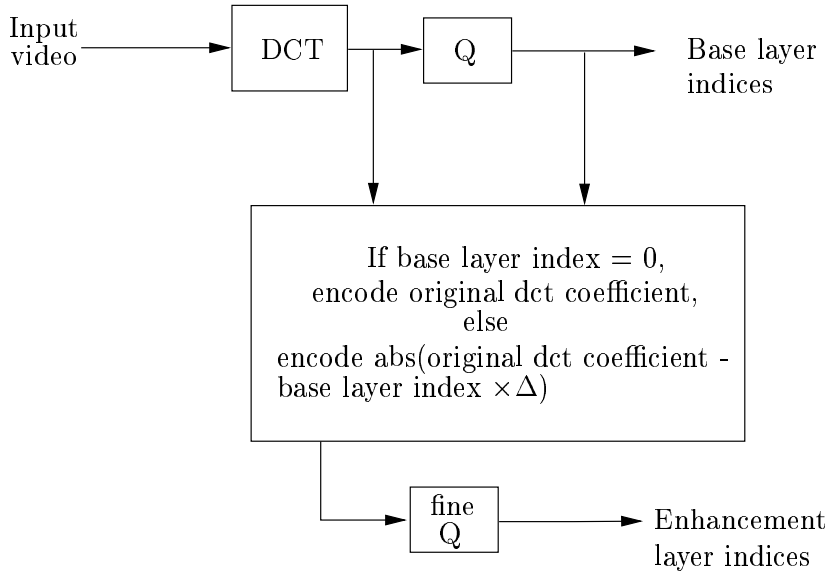

Figure 3.4: Block diagram of the new method used for SNR scalability.

In the existing method, the coding error between the base layer reconstruction and the original picture was computed in the pixel domain, then the DCT taken and fine quantization done. In the new method, the error is computed in the DCT domain, then quantized and transmitted. In this case the error is not simply a difference, but based on the quantization indices generated by the base layer encoder. At the receiver, the base layer indices and the enhancement layer indices are combined to give the final enhanced picture. As in

the existing method for SNR scalability, let Δ_{refine}, the refinement step-size be equal to $\Delta/2$. Comparing figure 3.4 to figure 2.7, it can be seen that the new enhancement method is computationally simpler since the DCT is taken once for each frame including enhancement.

P layer enhancement is similar to that of the existing method, but uses the DCT difference according to the new enhancement method.

3.4 Using TCQ for the Enhancement

This section discusses the use of trellis coded quantization for the enhancement coding rather than uniform scalar quantization. The base layer is coded using SQ and the enhancement using TCQ. The details of the SQ-TCQ encoder for the new method are given, followed by the results achieved after testing on memoryless data.

3.4.1 The SQ-TCQ Codec

Figure 3.5 shows the SQ-TCQ encoder block diagram. The base layer coder is a uniform threshold SQ with deadzone as in H.263+ Intra quantization.

The TCQ enhancement coders use an eight-state trellis, shown in Figure 3.6 as in [10]. The TCQ0 enhancement coder uses a symmetric codebook with two zero-valued reproduction points. The TCQ1 enhancement coder uses a one-sided codebook with a single zero-valued reproduction point. The quantized indices are arithmetic coded before trans-

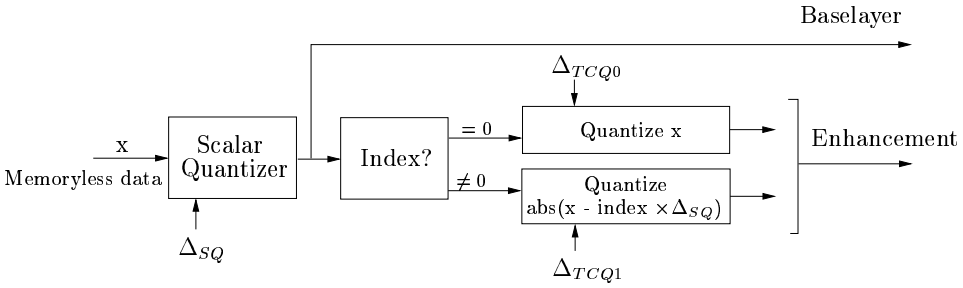

Figure 3.5: SQ-TCQ encoder for the new method.

mission.

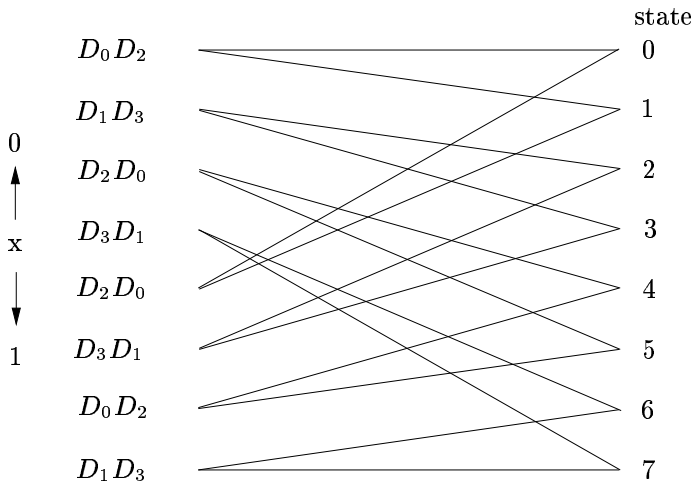

Figure 3.6: 8-state trellis with subset labels.

The reproduction alphabet is partitioned into four subsets(D_0, D_1, D_2, D_3). The encoders use as many uniformly spaced TCQ symbols as is necessary to cover range $[-\Delta_{SQ}, \Delta_{SQ}]$ for $(y = 0)$ and $[0, \Delta_{SQ}]$ for $(y \neq 0)$. The two TCQ codebooks are shown in Figure 3.7. The four subsets are combined to give two supersets as follows.

$$A_0 = D_0 \cup D_2, A_1 = D_1 \cup D_3 \tag{3.5}$$

The $(y = 0)$ coder uses the same arithmetic code model for both the supersets, and the $(y \neq 0)$ coder uses two models to account for the asymmetric codebook.

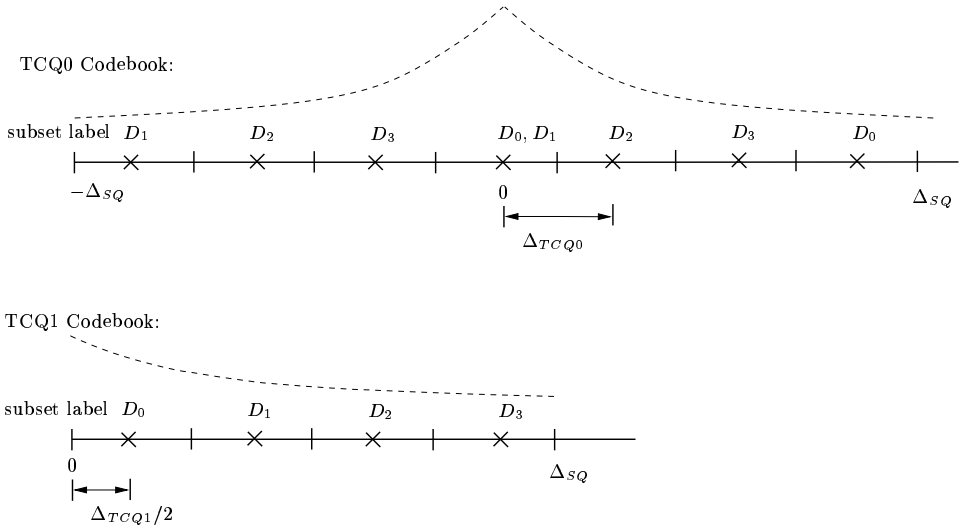

Figure 3.7: TCQ codebooks for $(y = 0)$ and $(y \neq 0)$ respectively.

Figure 3.8 shows the decoder. TCQ0data and TCQ1data are the decoded values from the respective TCQ decoder.

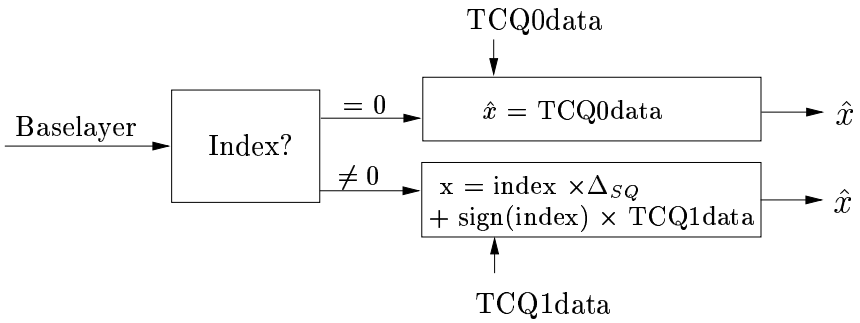

Figure 3.8: Decoder for the new method.

3.4.2 Results on Memoryless data

In the tests on memoryless data, Δ_{TCQ0} and Δ_{TCQ1} were both set to $\Delta_{SQ}/4$. Tests were performed on two cases of peaked source densities, namely Laplacian, and generalized Gaussian with $\nu = 0.5$. In each case 100,000 source samples were used. The three parameters are Δ_{SQ}, Δ_{TCQ0} and Δ_{TCQ1}. For an overall target rate R of 4.8 bits/sample, the SQ encoder was allocated R_{SQ} and the remainder of the rate shared between the two TCQ coders. It was found that if the TCQ was allocated sufficient combined rate, say, $R_{TCQ} \geq 1.5$ bits/sample, then the choice of $\Delta_{TCQ0} \approx \Delta_{TCQ1}$ gives the best SNR performance. The SNR versus bit rate results for zero mean, unit variance Laplacian data are shown in Figure 3.9.

The signal-to-noise ratio (SNR) is defined as

$$SNR = 10\log(\frac{VAR}{MSE})dB \tag{3.6}$$

The SQ-TCQ scalable codec is compared to a non-scalable arithmetic coded TCQ (ACTCQ) codec (with zero deadzone) and to a non-scalable entropy constrained scalar quantizer (ECSQ) codec both with and without deadzone. The SQ-TCQ scalable codec uses deadzone in the base layer only. The plot shows that as sufficient rate is allocated to the enhancement (TCQ), the SQ-TCQ codec performs better than an ECSQ at the same overall rate. As the R_{SQ} is reduced to about 0.8 bits/sample, the SQ-TCQ performance gets close to that of the non-scalable ACTCQ codec. For a particular application, one will

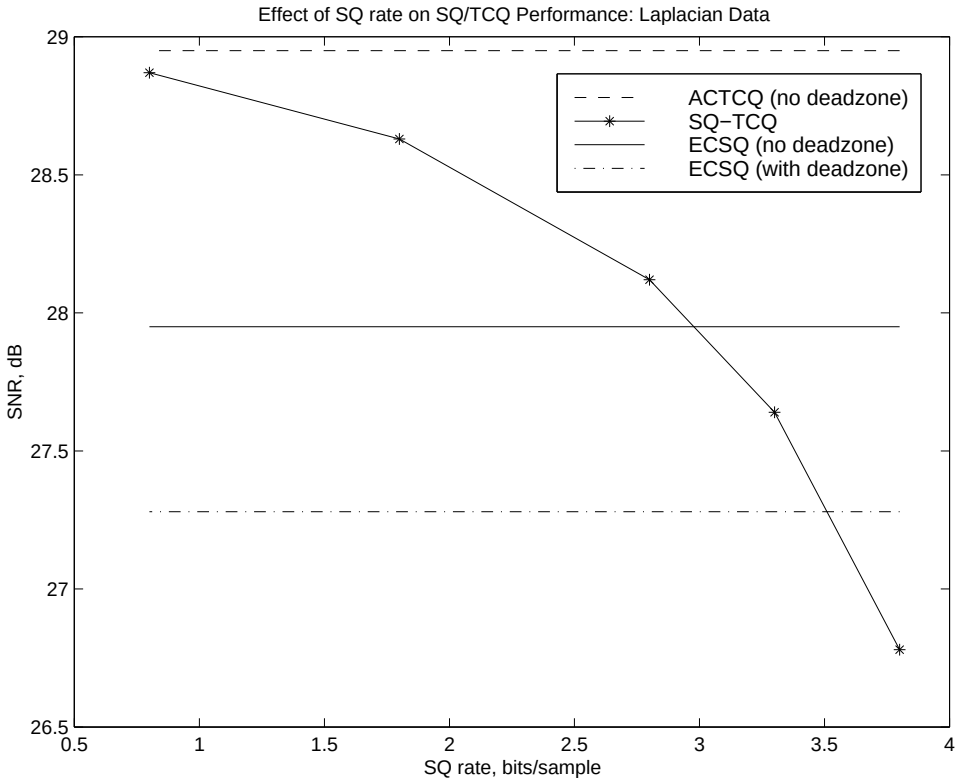

Figure 3.9: In the SQ-TCQ results, the total rate of 4.8 bits/sample is shared between the SQ and TCQ.

compromise between the quality of the base layer description and the SNR improvement in the enhancement layer.

The Shannon lower bound to the rate-distortion function is,

$$SNR_{SLB} = 10\log(\frac{\pi}{e}) + 6.02R = 29.52dB \tag{3.7}$$

where R takes the value 4.8 bits/sample.

Figure 3.10 shows the results for the scalable SQ-TCQ encoder compared to the non-scalable encoders for an overall rate of 2 bits/sample. Again Laplacian data is used. In this

case, the scalable coder performance exceeds the non-scalable ECSQ coder (with dead-zone) at an enhancement rate of about 1.1 bits/sample.

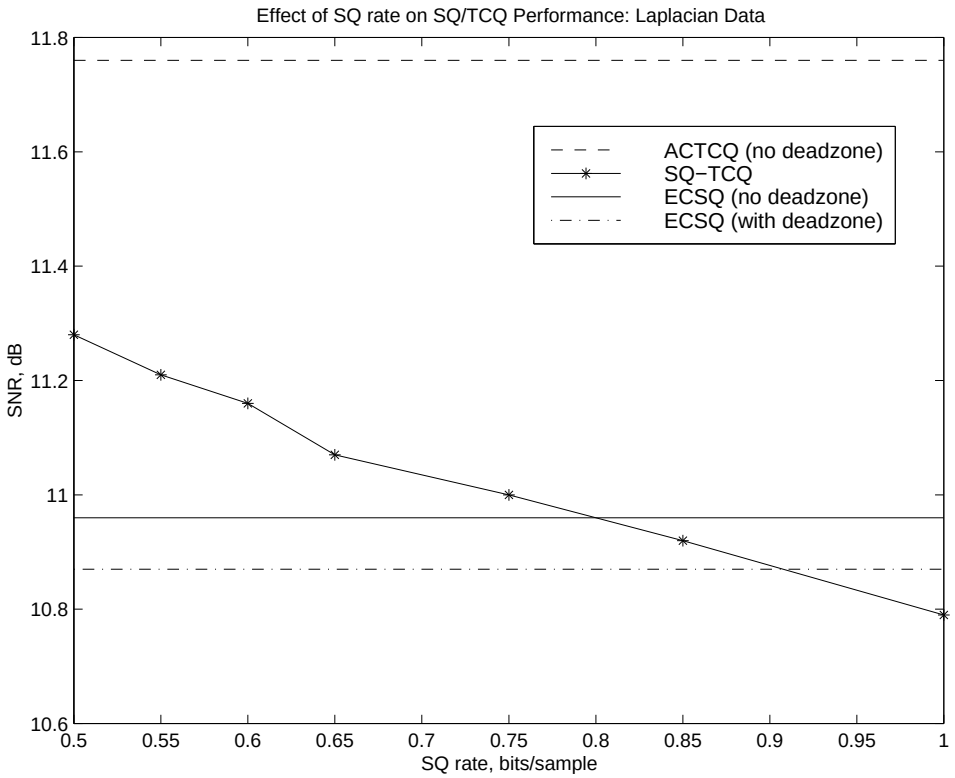

Figure 3.10: In the SQ-TCQ results, the total rate of 2.0 bits/sample is shared between the SQ and TCQ.

Chapter 4

Results of the New Method applied to H.263+

4.1 Introduction

This chapter presents the results of applying the new method of enhancement coding to the H.263+ SNR scalability mode. In the work done, both the base layer and the enhancement layer utilize uniform scalar quantization. Interesting points of discussion are included.

4.2 SQ-SQ Codec Results

To evaluate SNR scalability, three bitstreams were generated for the Carphone sequence. The first is the non-scalable description. The other two are the scalable descriptions for the existing and the new enhancement methods. For the scalable encoding, the scalar quantization step size is fixed for all pictures at the base layer. The enhancement layer step size is also fixed, equal to half that of the base layer. The non-scalable encoder uses the same fixed step size as that of the enhancement layer in the scalable encoder. The bit rate of the

enhancement layer is defined as the total bit rate for both the base and the enhancement layers. Figure 4.1 shows the results of the bit rate versus luminance PSNR for the enhancement layer of the new method compared to the existing method and to the non-scalable description for the I frame. The peak signal-to-noise ratio (PSNR) is defined as

$$PSNR = 10\log(\frac{(255)^2}{MSE})dB \tag{4.1}$$

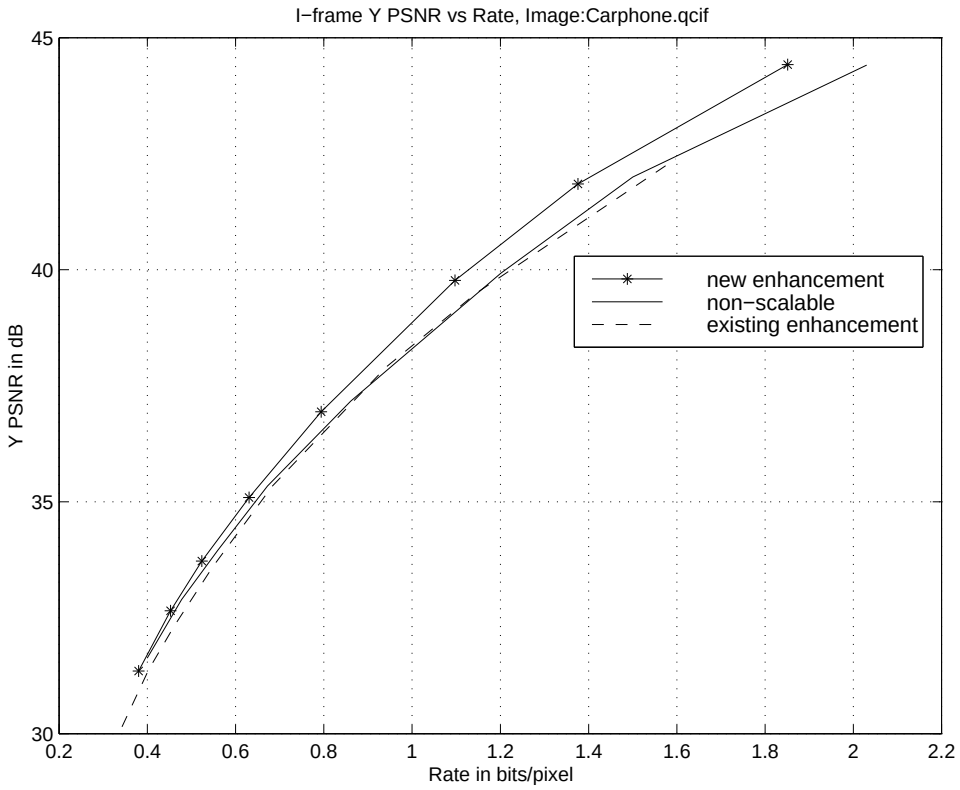

Figure 4.1: I frame results for SNR scalability for the Carphone sequence.

In this case, the performance of the new scalable codec exceeds both that of the existing enhancement method as well as that of the non-scalable codec. The explanation for why

the new method performs better than the non-scalable codec is that since the non-scalable description uses a small step-size directly on large DCT coefficients, there is a greater likelihood of run-level sequences that do not occur in the VLC frequency tables. In this case 22 bits are used to code the run-level sequence which results in an increased overall bit rate. The scalable coding, on the other hand, uses a larger step-size on the input data, then a smaller step-size on the refinement data. From a statistical check on coding the I frame of the Carphone sequence using a final step-size of 4, the scalable coder uses the 22-bit escape sequence less than half as many times (in the base layer only), as the non-scalable coder. It can be concluded, therefore, that the H.263+ entropy coding (variable length coding using a combined Huffman and run-length table) is not necessarily well-tuned to coding of I frames.

Figure 4.2 shows the visual results, comparing the decoded I frame of the Carphone sequence using the new and existing enhancement coding methods. The PSNR of the new method enhancement picture is 1.40 dB higher than the existing method enhancement, but the rate of the new method is also higher. This is due to more DCT coefficients being enhanced by a non-zero value compared to the existing method. From figure 4.1, comparing the two methods at a rate of 1 bit/pixel shows a PSNR improvement of about 0.6 dB using the new method. This improvement increases with increasing rate.

Figure 4.3 shows the base layer description in the scalable coding, and the non-scalable coding picture. In the case of bandwidth restriction for example, the base layer in the scalable coding provides at least a very recognizable picture at the receiver.

a) New method enhancement

b) Existing method enhancement

Figure 4.2: Comparing the EI frames, using a step-size of 20. For the new method, R = 0.52 bits/pixel, PSNR = 33.72 dB, for the existing method, R = 0.46 bpp, PSNR = 32.32 dB.

Figure 4.4 shows the results for the P frame encoding, averaged over 16 encoded P frames. Figure 4.5 shows the overall results after encoding one I and sixteen P frames. Consistent with [2] the results for the overall encoding show about a 1-2 dB drop in PSNR in the scalable encoding compared to the non-scalable. This drop is reduced in the new scalable encoding technique. The reason that the new refinement method does not yield performance better than non-scalable coding at the higher bit rate, for P frames, and the

a) Base layer I frame

b) Non-scalable I frame

Figure 4.3: I frame for the base layer in the scalable coding using a step-size of 40 and the non-scalable I frame using a step-size of 20. For the base layer, R = 0.30 bits/pixel, PSNR = 30.05 dB, for the non-scalable description, R = 0.56 bits/pixel, PSNR = 33.96 dB.

combination of I and (many) P frames is that when refinement is used, the base layer representation has significant MSE, so that the motion-compensated prediction forms differences that have more information to code, than when a higher bit rate is used with no enhancement.

In the new method, all enhancement DCT coefficients are treated as Intra in the quantization. This is different from the existing method which treat enhancement coefficients

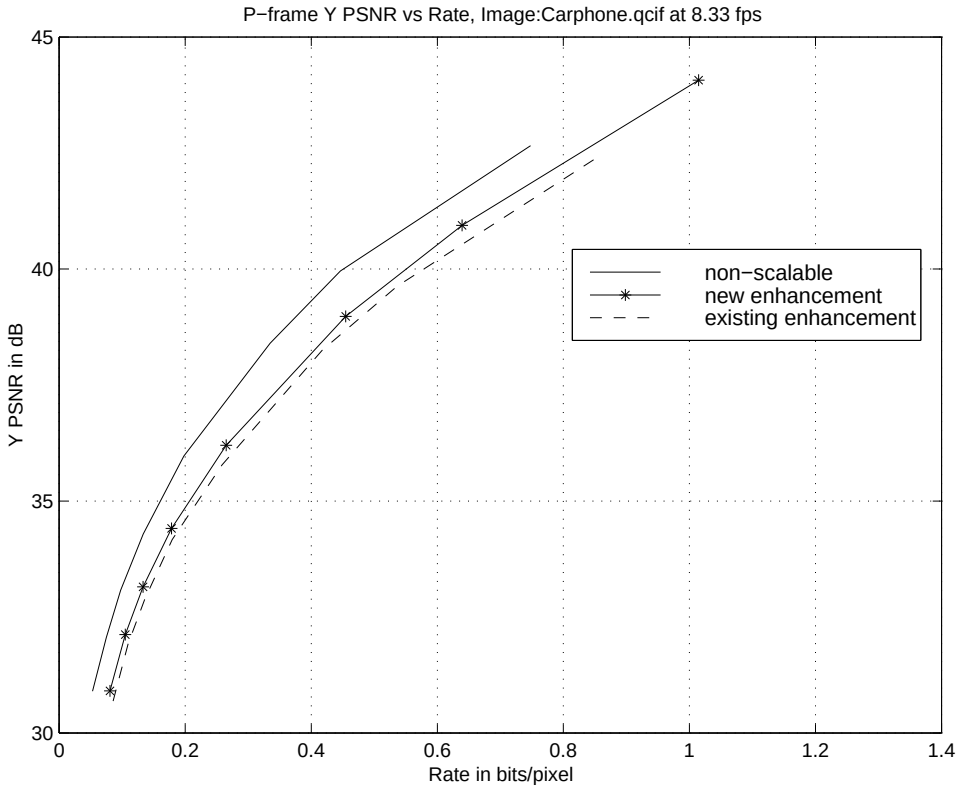

Figure 4.4: Average P frame results for SNR scalability for the Carphone sequence.

as Inter. In other words, the new method does not use an increased deadzone in the quantization of the enhancement coefficients. This improved the new codec performance significantly. Also, in the reconstruction, rules 2.6, 2.7 were not followed. Straight-forward midpoint reconstruction was used.

In the entropy encoding only one VLC table was used to encode the quantized data in the new enhancement method. The last non-zero level was not coded using a second VLC table, like H.263+. This reduced the encoding rate slightly.

Figure 4.6 shows the I frame results for the Foreman sequence. This result is similar to that of the Carphone sequence above. Figure 4.7 shows the overall results for the Foreman

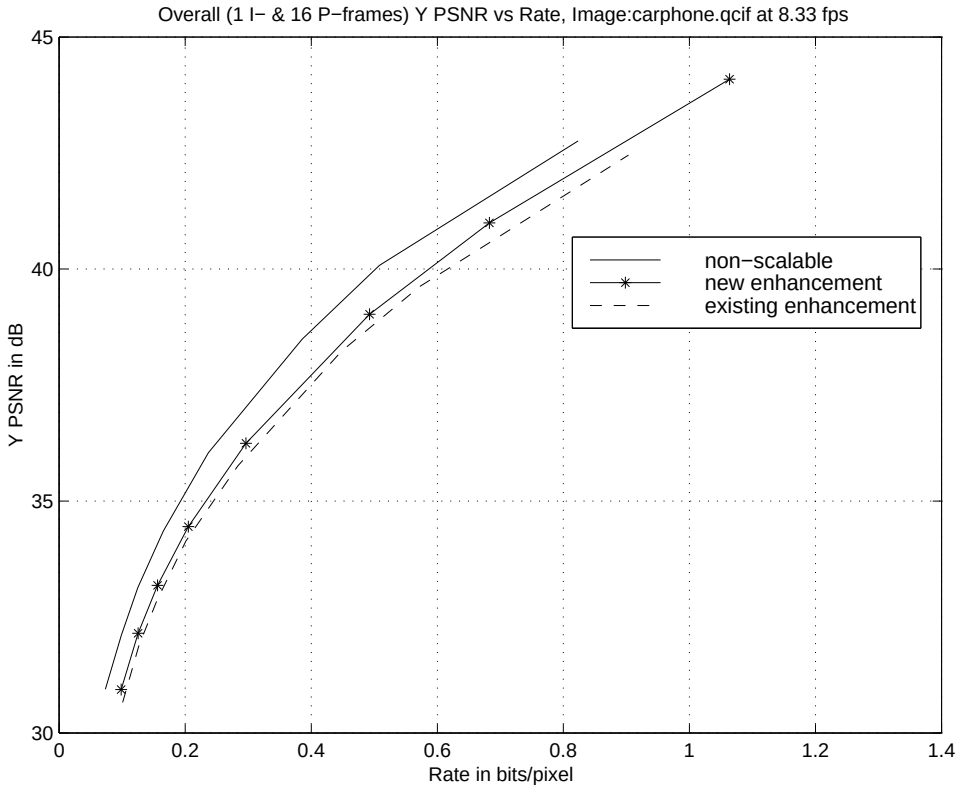

Figure 4.5: Overall results for SNR scalability for the Carphone sequence.

sequence. In video sequences that contain high motion and high camera motion, such as the Foreman sequence, occluded macroblocks occur more frequently. In coding a P frame, an occluded macroblock must be Intra coded. Since the new method gives good results for Intra coding, coding sequences such as Foreman using the new enhancement method promises to give additional improvement over the existing method. However, in the work done occluded macroblocks were not coded using the new method.

Figure 4.8 shows the I frame results for the Miss America sequence. The Miss America sequence uses lower rate. For this case the new enhancement method performs better in PSNR than the non-scalable coding only for rates above about 0.45 bits/pixel.

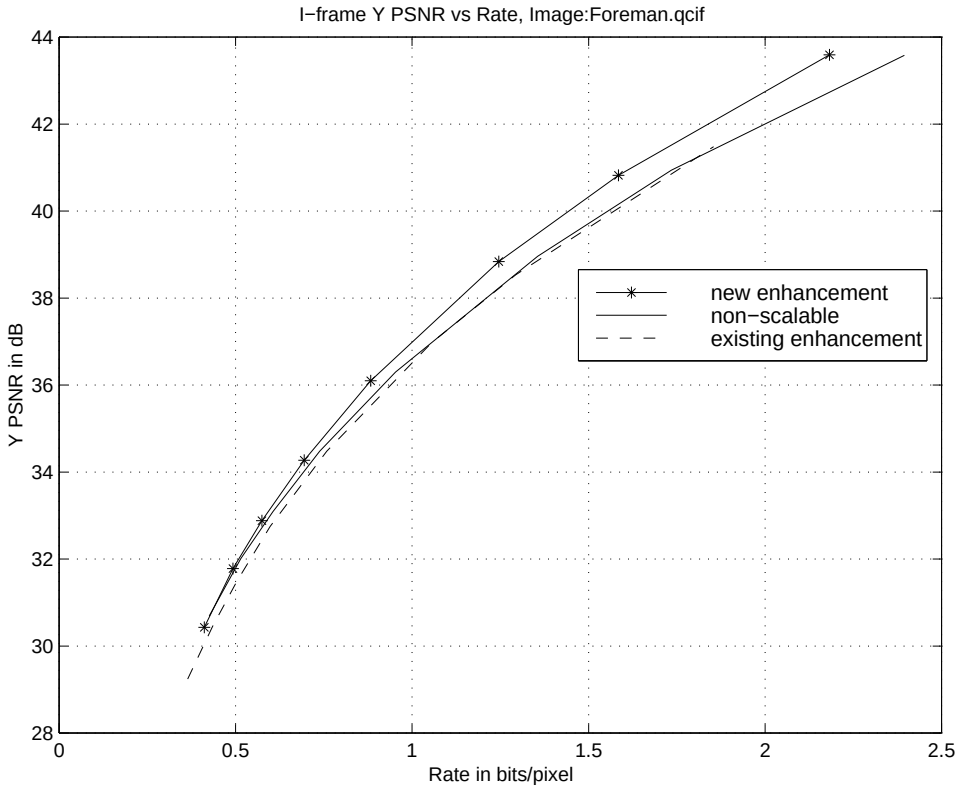

Figure 4.6: I frame results for SNR scalability for the Foreman sequence.

Figure 4.9 shows the overall results for the Miss America sequence. These results are also similar to that of the Carphone sequence.

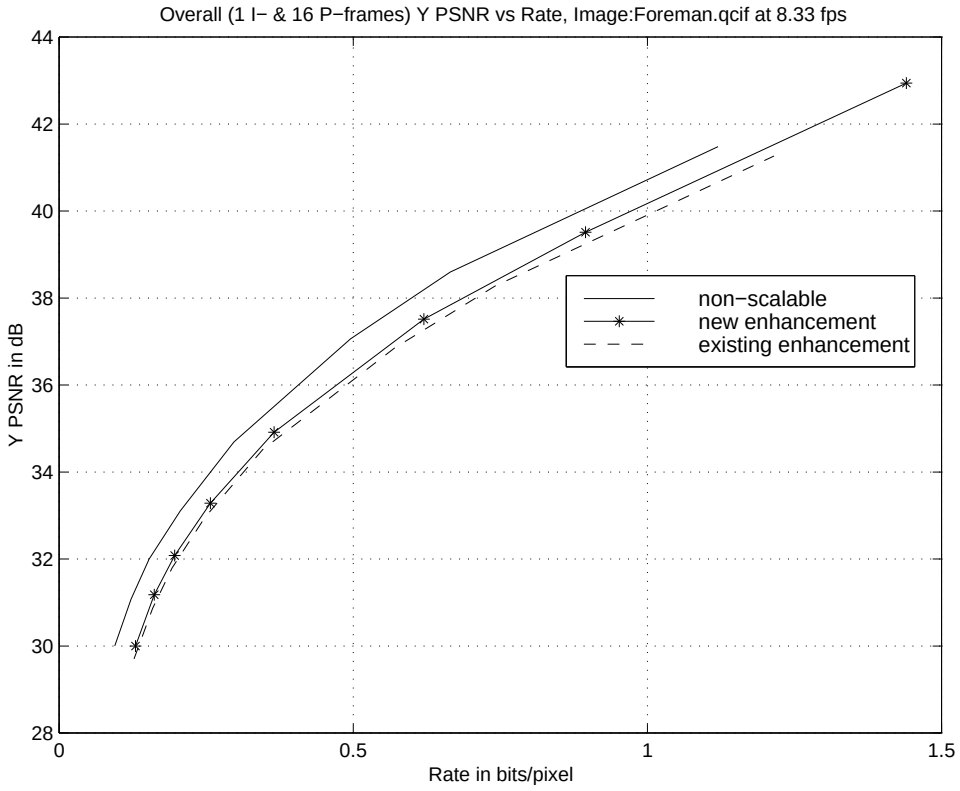

Figure 4.7: Overall results for SNR scalability for the Foreman sequence.

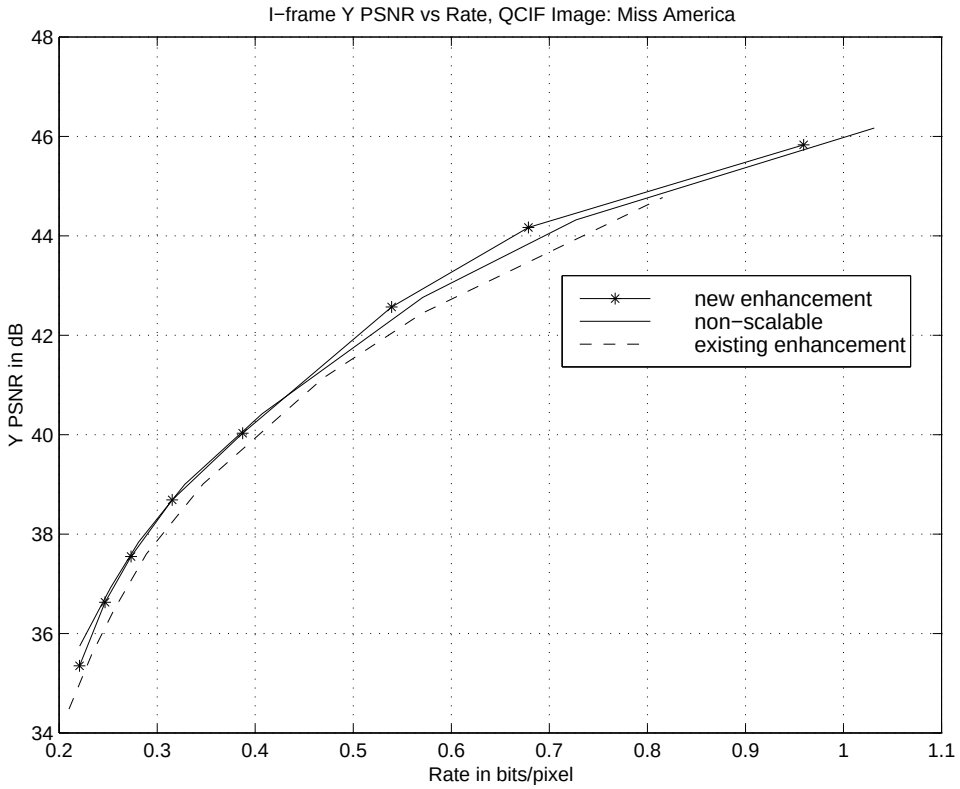

Figure 4.8: I frame results for SNR scalability for the Miss America sequence.

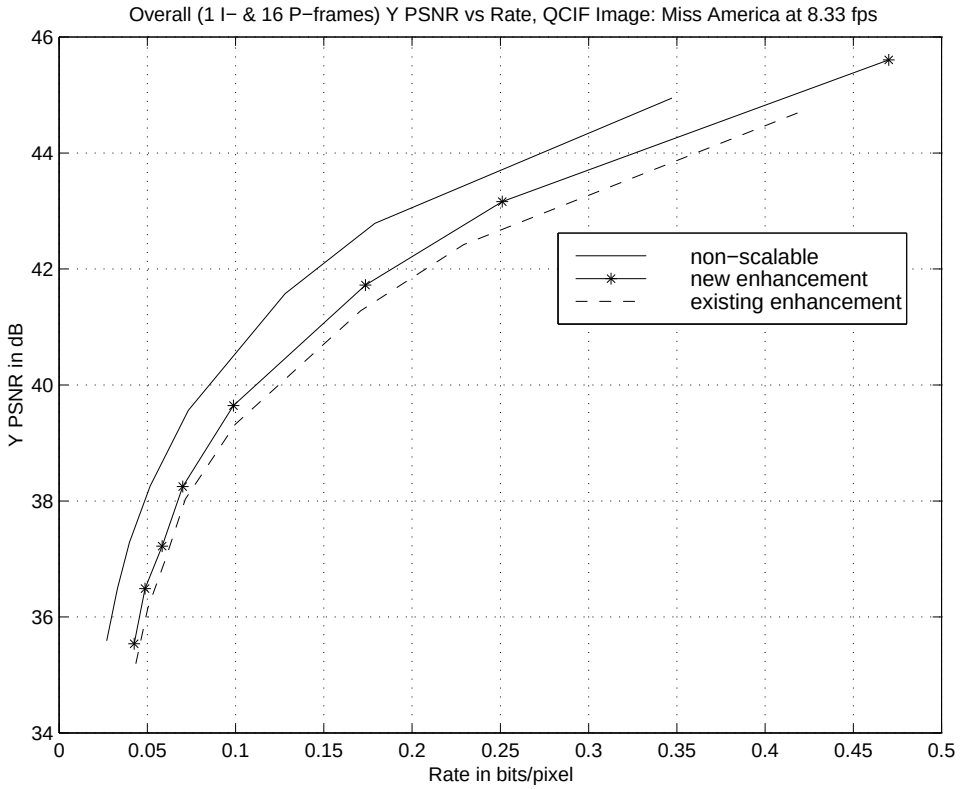

Figure 4.9: Overall results for SNR scalability for the Miss America sequence.

Chapter 5

Conclusion

5.1 Summary of Thesis Contribution

In this work, a new method for the encoding of enhancement frames has been proposed. Results obtained after applying the new SQ-SQ method to H.263+ SNR scalability coding show that a significant improvement in SNR is obtained over the existing scalability method. It was found that the use of an increased deadzone, as proposed by the existing standard, results in a sub-optimal enhancement. The new method maintains a fixed quantization rule, in both the base and enhancement layers. The new enhancement method is also less complex than the existing enhancement method.

Based on the results obtained in using SQ for the base layer encoding and TCQ for the enhancement layer, on memoryless data, it is expected that using a SQ-TCQ codec in H.263+ SNR scalability coding would give improved performance.

5.2 Areas of Further Research

There are several areas of research that have become unveiled by this investigation. One is a careful study of the use of deadzone in coding enhancement pictures. The parameters of interest will be the deadzone value, bit rate and picture quality.

An interesting investigation would be to construct better VLC codes both for the scalable and non-scalable encoders. The use of the H.263+ syntax-based arithmetic coding option can also be surveyed in the context of the new scalable coding method.

Another interesting area is the investigation of the overall benefit of applying the new method of enhancement while operating in a data-lossy environment, with the use of error-correcting or error-concealing techniques.

Finally, the new enhancement method can be applied to the spatial scalability option in H.263+.

Bibliography

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, New Jersey: Prentice Hall, 1971.

[2] G. Côté, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video Coding at Low Bit Rates.", *IEEE Transactions on Circuit and systems for Video Technology*, vol. 8, pp. 849-866, Nov. 1998.

[3] T. R. Fischer and M. Wang, “Entropy-constrained trellis coded quantization,” *IEEE Transactions on Information Theory*, vol. 38, pp. 415-426, Mar. 1992.

[4] L. E. Franks, *Signal Theory*, Dowden & Culver, Stroudsburg, PA, 1981.

[5] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[6] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, 1992.

[7] ITU - Telecommunication Standardization Sector, “Video Coding for Low Bitrate Communication,” ITU-T Recommendation H.263, January 1998.

[8] ITU - Telecommunication Standardization Sector, “Test Model 11,” February 1999.

[9] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Englewood Cliffs, New Jersey: Prentice Hall, 1984.

[10] R. L. Joshi, V. J. Crump, T. R. Fischer, “Image subband coding using arithmetic coded trellis coded quantization,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 515-523, Dec. 1995.

[11] M. W. Marcellin and T. R. Fischer, “Trellis coded quantization of memoryless and Gauss-Markov sources,” *IEEE Transactions of Communications*, vol. 38, pp. 82-93, Jan. 1990.

[12] M. W. Marcellin, “On entropy-constrained trellis coded quantization,” *IEEE Transactions of Communications*, vol. 42, pp. 14-16, Jan. 1994.

[13] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Inc, 1991.

[14] K. R. Rao and Z. S. Bojkovic, *Packet Video Communications over ATM Networks*, Prentice Hall PTR, Upper Saddle River, NJ, 2000.

[15] S. J. Solari, *Digital Video and Audio Compression*, McGraw-Hill, New York, NY, 1997.

[16] University of British Columbia, Signal Processing and Multimedia Group, “H.263+ Library Software Codec Simulator Version 0.2” January 2000.

[17] I. H. Witten, R. M. Neal, J. G. Cleary “Arithmetic Coding for Data Compression,” *Communications of the ACM*, vol. 30, pp. 520-540, June 1987.