A New Transform Domain Vector Quantization Technique for Image Data Compression in an Asynchronous Transfer Mode Network.

by

Peter P. Polit, student member IEEE and Nasser M. Nasrabadi, member IEEE.

Electrical Engineering Department
Worcester Polytechnic Institute
Worcester, MA 01609

1. INTRODUCTION

Picture transmission is a very effective method of conveying information for a large number of applications. Because of the large amount data and channel losses present in digital systems, data compression techniques are developed for reducing transmission bit rate and diminishing the noticeable effects of transmission errors.

The issue of concealing the transmission errors becomes more pressing in a packet switching environment, where one deals with a loss of a cluster of bits (a cell). In particular, the effect of cell loss needs to be carefully reviewed in light of the recently adopted standards for Asynchronous Transfer Mode (ATM) multiplexing technique networks that use Variable Bit Rate (VBR) transmission. The CCITT Study Group VIII adopted a 5 octet header and 48 octet information field structure [1a]. In this environment, cell loss (at rates expected to be between $10^{-3}$ and $10^{-5}$) can not be easily compensated for. Unlike bit errors, cell loss can not be easily improved by forward error correction [1]. Cell loss detection and correction algorithms have been successfully designed [2],[3], but their complexity has remained relatively high.

Susceptibility to cell loss is what has made most image coding techniques less attractive in the ATM environment. In the past, numerous otherwise effective picture coding techniques have been developed including Vector Quantization (VQ) [4-5], which is an extremely powerful compression technique, but suffered significantly in the presence of packet loss. To increase VQ’s efficiency, diverse attempts have been made to combine the transform domain coding with VQ [1],[6-7]. Although this increases the encoding technique’s resistance to channel noise, the effect of cell loss on the encoded images remains a major problem.

The cell loss problem has persisted for all recursive coding techniques as well, including differential pulse coding modulation (DPCM). In both spatial and temporal DPCM, a loss of one element significantly affects the next one and erodes the coded image, and the error propagates until an element refresh cycle takes place. Although a number of error concealment techniques to limit error propagation were developed and incorporated into a variety of two layer hybrid DCT DPCM coding approaches [8-10], DPCM remains difficult to implement in an ATM environment.

The negative effect of cell loss was reduced with introduction of sub-band coding techniques [11-12]. These techniques subdivide the signals into bands of more and less important information. The base band needs to be transmitted with virtually no errors, but a large cell loss can be tolerated in the less important bands without sacrificing the reconstructed image quality. These techniques generally require a significant priority layer within the network.
Other transform domain techniques were developed incorporating the Lapped Orthogonal Transform in order to reduce the blocking effects [13], as well as channel noise. With the blocking effect reduced, the block size can be increased while retaining the same transmission quality. With the size of the block increased, the number of blocks, together with the number of DC components, can be decreased. However, the amount of high priority information that needed to be transmitted still remains considerable.

This paper presents a new coding technique suitable for an ATM network. It consists of VQ combined with Discrete Cosine Transform (DCT). This technique requires a small or no priority layer, is computationally inexpensive to implement, and is robust to packet loss. This coding strategy was extensively tested and found to be suitable for an ATM environment.

2. AN OVERVIEW OF APPLICABLE IMAGE CODING TECHNIQUES

The following is a brief overview of the coding techniques that served as the basis for the work performed in this project. These techniques were selected on the basis of their performance in data compression and applicability to lossy transmission environment.

2.1 Vector Quantization

Vector Quantization (VQ) is the most computationally demanding technique implemented in this project as one of the image coding steps. VQ capitalizes on the Shannon rate-distortion theory, which states that a better quantizer performance is always in theory achievable by coding vectors instead of scalars, even though the data source is memoryless [5]. Vector Quantization has been extensively examined over the years and numerous approaches have been developed.

A vector quantizer uses a finite set of vectors as its mapping space. This vector set is referred to as the codebook. A complete system consists of an encoder which assigns to each input vector \( x = (x_1, x_2, \ldots, x_n) \) a symbol \( \gamma(x) \) (called the address) in the mapping space, and a decoder that associates the received address with the proper mapping space (also called codebook) element [4]. If the codebook consists of \( M \) vectors, then the rate of the quantizer is defined as \( R = \log_2 M \) bits per vector, and \( r = R/k \) bits per symbol, making this a compression technique with great transmission rate reduction potential. The reproduction mapping set need not be the same as the input codebook [4].

In order to successfully quantize an image, one has to define a distortion measure to evaluate the image quantization error. A quantizer is said to be optimal when the distortion is minimized for a given image. One popular distortion measure is the mean square error defined as:

\[
    d(x, \hat{x}) = |x - \hat{x}|^2 = \sum_{i=0}^{n-1} (x_i - \hat{x}_i)^2
\]

where the error is actually the Euclidean distance between the reproduction vectors and image vectors. A simple input-dependent weighting was showed to improve the correlation between this distortion measure and the observed image quality [4-5].

An effective non-parametric quantizer codebook design proposed by Linde, Buzo, and Gray and commonly referred to as the LBG algorithm was implemented in this work [14]. This algorithm uses an initial codebook and a training set of images to arrive at a codebook of unknown statistical distribution as described in [14]. The mean square error was used as the distortion measure in the design process and produced acceptable results.
2.2 Discrete Cosine Transform (DCT)

Transform coding in general enables the data to be represented by a restricted set of transform coefficients that can be discarded, transmitted, or stored for subsequent inverse transformation and image reconstruction [15]. The most efficient transform is one that packs the maximum amount of signal energy into the least number of coefficients, so that most of the low energy coefficients can be discarded. A transform that performs extremely well in energy compacting of highly correlated sources, and is data independent, is the Discrete Cosine Transform (DCT). The DCT is a real transform that can be implemented using a fast algorithm. This work uses the even version of the one-dimensional transform defined as:

\[ C(k) = \alpha(k) \sum_{n=0}^{N-1} u(n) \cos \left( \frac{\pi(2n+1)k}{2N} \right), 0 \leq k \leq N - 1 \]  
\[ \alpha(k) = \begin{cases} 1 & k = 0 \\ \frac{1}{\sqrt{N}} & 1 \leq k \leq N - 1 \end{cases} \]

The DCT is real and orthogonal and it is a fast transform. Since DCT decorrelates the data effectively, it is very appealing for use in the ATM environment, as a loss of one coefficient does not impair the next one directly. That is, the next frequency coefficient can be recovered exactly, but the spatial domain image still suffers quality deterioration though to a smaller extent and over a wider area than it would in case of a spatial loss.

2.3 DCT VQ

DCT VQ techniques are very promising when it comes to reducing transmission bit rates as well as increasing resistance to channel noise. However, those techniques are generally very susceptible to cell loss, as losing any vectors within any DCT block significantly deteriorates the block’s visual quality in spatial domain. To avoid the effect of cell loss but to exploit the advantages of the DCT, a hybrid coding technique was introduced in this paper that prevented multi-frequency component loss in the transform domain.

3. ONE-DIMENSIONAL TRANSFORM DOMAIN INTERLINE VQ

A new technique of combining simple VQ with one dimensional DCT is proposed in this section. The technique, One Dimensional Transform Domain Interline VQ, is computationally inexpensive, reduces the required bit rate significantly, and is robust to packet loss inherent to ATM networks. In this section of the paper we present the
proposed technique applied to coding still image frames. This is extremely useful in some applications, including initial frame transmission for moving sequence coding process.

3.1 Codebook design

The proposed coding technique clustered the corresponding harmonic coefficients into groups of sixteen. Figure 2 illustrates the one dimensional DCT applied to an image, and the coefficient clustering method. An AC coefficient in one line was clustered with corresponding AC coefficients from 15 other lines to form a 16 coefficient cluster (vector). The authors found that increasing the vector size beyond 16 coefficients, without increasing the codebook size, resulted in significant coded image quality degradation (highly perceptible image blocking). For vectors smaller than 16 elements, image quality improved slightly, while increasing the required transmission rate significantly. The marginal utility of decreasing the vector size was therefore low. These statements were experimentally proved for 512x512 image frames, and confirmed on motion compensated 352x288 moving image frames.

Due to the highly varying statistical distribution of various frequency coefficients for each line of a given image, it was highly inefficient to create only a single codebook for the entire image. This was the case because of the fact that the LBG algorithm used for codebook generation requires a stationary source. Stationarity could not be assumed over the entire transformed image, but could be approximated by dividing the transformed image into nearly stationary zones that were determined experimentally. Therefore, a flexible codebook training field method was adopted in this paper. Figure 1 illustrates this procedure. All the transform domain coefficient vectors within a given training field were assumed stationary. Also, the training field width was experimentally determined for each image region to produce satisfactory reproduced image quality. The high quality of reconstructed images reconfirmed the assumption of vector stationarity within each image training strip. The DC coefficients were not included in the clustering process. They were separately scalar quantized. This approach was necessitated by the fact that the human visual system is most sensitive to brightness variations, and these variations would arise from clustering the DC coefficients.

The total number of created codebooks was eighteen. This number was experimentally determined to be an optimal compromise between codebook storage requirements and the coded image quality. Each codebook consisted of 512 clusters of 16 AC coefficients extracted from 16 neighboring image lines. Each coefficient cluster contained coefficients belonging to exactly the same frequency on respective transformed image lines.

3.1.1 Initial codebook design

The initial codebook was developed using a training sequence of images. The images were selected to contain a wide variety of spectral information ranging from a rapidly changing landscape with high detail content to a relatively uniform facial portrait. The first 512 vectors (normalized to approach zero mean and unit variance) were chosen in each area of interest (Fig. 2) in the training sequence. The distance between every two vectors was incrementally increased during the design process. When the required minimum distance grew large enough to make it impossible to select more than 512 vectors from the training set, the design process was stopped. The resulting 512 element initial codebook spanned the entire set of training sequence vectors within the area of interest, meeting a minimum inter-vector distance requirement.
3.1.2 Final codebook design

The final quantizer design was performed by applying the LBG algorithm to image strips of varying width, as pictured in Fig. 2. The initial codebooks served as the starting point for the LBG algorithm. The LBG algorithm could be implemented, because the source signal (transformed video frame) was assumed to be stationary in a given training frequency band, as described in section 3.1. A different codebook was created for each image region. It was found that it is necessary to make the low frequency (high energy region) training strips very narrow, (2 coefficients at a time), in order to retain coded image quality above 35dB. The low energy (high frequency) coefficients were not nearly as demanding, and the training strip size 75 coefficients wide performed quite well. Unpleasant, grainy images resulted from high frequency training strips wider than 75 coefficients. For low frequency strips wider than two coefficients, disturbing image “streaking” resulted, as the vectors were no longer accurately represented by a codebook of 512 codewords.

The LBG quantizer used the convenient squared error distortion measure defined as:

\[ d(x, \hat{x}) = \sum_{i=0}^{k-1} |x_i - \hat{x}_i|^2. \]

The algorithm was used on an unknown statistical distribution series of transformed images, and a quantizer was designed as described in [14]. Once the final codebook was obtained, the experimental image was coded in respective strips of coefficients using the optimum nearest neighbor rule. The codebook design process was performed on mean and variance normalized coefficient clusters. The clusters were normalized with respect to the algebraic mean of the element variances, i.e.:

\[ \sigma_{\text{mean}}^2 = \frac{1}{16} \sum_{i=0}^{15} \sigma_i^2. \]

Since the question of applicability of the algebraic mean of variances has been raised in the past [7], the geometric mean was applied in some simulations, i.e.:

\[ \sigma_{\text{mean}} = \sqrt{\frac{\prod_{i=0}^{15} \sigma_i^2}{16}}. \]

Although it can be shown that the best degree of deviance indicator for a given vector is in fact the ratio of arithmetic and geometric variances, \( \frac{\sigma_{\text{mean}}}{\sigma_{\text{mean}}^2} \), in this particular application, no significant change in coded image quality was experienced. The vector mean was computed to be the algebraic mean of the elements, and the normalized data approached zero mean and unit variance distribution.

Finally, when coding, a full search procedure was implemented for the appropriate codebook in a given image region. Each ATM cell was packed with coefficients belonging to exactly one frequency. In case of a loss, only one frequency coefficient is missing on a given line, and the image can be successfully reconstructed.

3.2 Bit rate, storage requirements, and coded quality

In the final coded images, each group of 16 coefficients was described by a 9 bit address, resulting from the codebook size selection of 512. With the DC coefficients scalar quantized at 8 bits, the resulting bit rate was approximately 0.58 bits per pixel (bpp).
A total of 18 codebooks were created, and each contained 512 clusters of 16 floating point numbers represented by four bytes of data. Figure 3(a) is the original experimental image LENA, and Fig. 3(b) depicts the coded version with no packet loss present. The resulting coded image PSNR was 42.73 dB.

The subjective quality of the coded image was good. By numerous spectators it was judged to be between (1) and (2) on the impairment scale in section 3.3 of this paper. In conjunction with significant bit rate reduction and resistance to packet loss, this was found to be a satisfactory result.

3.3 Distortion measures

Both objective and subjective distortion measures were employed in evaluating the images. When searching for optimal codebook training technique, at least three spectators were asked to subjectively evaluate the resulting coded image. In all cases, the subjective score corresponded to the objective SNR rating. The peak signal-to-noise ratio (PSNR) was computed using the following:

\[
SNR = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \frac{(x-x)^2}{(N^2)(\text{number of gray levels})^2}, \text{where } N = \text{image size.} \quad (9)
\]

\[
SNR_{\text{dB}} = 10 \log_{10}(SNR) \quad (9a)
\]

This formula was used in quantitative evaluation of all of the images.

The subjective rating was conducted using an impairment scale borrowed from Jain [16]. The scale is depicted in Table 1.

<table>
<thead>
<tr>
<th>Impairment Scale</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not noticeable</td>
<td>(1)</td>
</tr>
<tr>
<td>Just noticeable</td>
<td>(2)</td>
</tr>
<tr>
<td>Definitely noticeable, slight impairment</td>
<td>(3)</td>
</tr>
<tr>
<td>Impairment, not objectionable</td>
<td>(4)</td>
</tr>
<tr>
<td>Somewhat objectionable</td>
<td>(5)</td>
</tr>
<tr>
<td>Definitely objectionable</td>
<td>(6)</td>
</tr>
<tr>
<td>Extremely objectionable</td>
<td>(7)</td>
</tr>
</tbody>
</table>

3.4 Packet loss

As the final part of the simulation, the effect of packet loss upon this technique was investigated. The simulated packet loss fell in four general categories. In category 1, 1% of all frequency coefficients was lost in a sample coded image LENA. This assumes no priority layers within the network. The result was very good, and the deterioration was barely noticeable. The subjective score fell just below (2) on the impairment scale, and the objective PSNR of the degraded image was recorded at 39.28 dB. The 1% packet loss rate was excessive, as cell loss rates are usually estimated at 10\(^{-4}\) to 10\(^{-5}\) for channel rates below 140 Mbits/s. The channel rate for this simulation was approximated at 4.6 Mbits/s. Figure 4(a) illustrates the packet loss map indicating the removed AC DCT components. The data voids were simply filled with zeros, and no effort was made to recover from packet loss by producing artificial data. Figure 4(b) illustrates the image reconstructed after a random 1% packet loss.
The second category was a simulation of a layered, congested network with 10% of the high frequency cells lost (above the 256th components). The network was assumed capable of sending the first half of all frequency coefficients intact. Here, the result again was satisfactory, with the subjective score at (2) and PSNR of 41.51 dB. The resulting image with the corresponding packet loss map is presented in Fig. 5. In the same manner 1% of all cells above the 256th AC frequency element were lost in another simulation. The resulting PSNR was 42.48 dB, and the quality drop was not visible.

The third category was the simulation of a network with only a limited priority layer. Here, the first 20 frequency components were protected. 10% of the entire image was missing, and the voids were randomly distributed among the transform coefficients above the first 20. The subjective rating of this image was placed at (3) on the impairment scale. The PSNR of the reconstructed image was at 30.81 dB. The resulting image and the packet loss map are presented in Fig. 6.

Category IV was the simulation of an extremely poor performance, congested network, where 10% of all packets were randomly lost. Here, the deterioration was significant, but the image frames were still easily recognizable. The subjective score fell between (5) and (6) on the impairment scale. The objective PSNR was at 25.95 dB. Figure 7 shows the resulting image.

Overall, the packet loss simulation experiments confirmed the applicability of the developed coding technique to the ATM environment. Tabulated below are the experimental results.

<table>
<thead>
<tr>
<th>Image loss category</th>
<th>Subjective rating</th>
<th>Objective rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>(2)</td>
<td>39.28 dB</td>
</tr>
<tr>
<td>II</td>
<td>(2)</td>
<td>41.51 dB</td>
</tr>
<tr>
<td>III</td>
<td>(3)</td>
<td>30.81 dB</td>
</tr>
<tr>
<td>IV</td>
<td>(5)-(6)</td>
<td>25.95 dB</td>
</tr>
</tbody>
</table>

3.5 Vector interleaving

In the cell loss simulation presented here, all ATM cells were packed with the same frequency coefficients. This resulted in the benefit of losing only one frequency coefficient per line in case of a lost cell. The drawback associated with this method was that 80 neighboring lines lost the same frequency coefficient. This resulted in perceptible reconstructed image zoning at high packet loss rates.

To alleviate the reconstructed image zoning problem, it is possible to pack ATM cells with clusters consisting of various frequency terms. A cell containing the first AC coefficient cluster from one group of 16 lines (a significant one), could for instance also contain a high frequency one (insignificant) from the next group. In case of a loss, the first 16 lines would suffer to a greater extent, but the second group would be reconstructed nearly perfectly. Although not raising the objective PSNR, this approach, labeled vector interleaving, would produce images more pleasing to the eye at high cell loss rates.
4. CONCLUSIONS

A new ATM oriented transform domain VQ technique was presented in this paper. This technique, One-Dimensional Transform Domain Interline VQ, shows resistance to packet loss effects, is simple to implement, and gives good compression rates. With the transmission bit rate under 4.6 Mbit/sec (30 frames/sec), the cell loss rates were overestimated at under 1%, and virtually no visible quality deterioration in the transmitted image resulted. At more realistic network cell loss rates of 10^−4, there was no quality deterioration at all in the image reconstructed at 0.58 bpp.

The computational costs of this technique can be estimated as low. This, combined with the technique’s performance, makes the technique a worthy candidate for system implementation.

REFERENCES


Fig. 1. DCT coefficient clustering.

Fig. 2. Flexible training fields.

Fig. 3. (a) The original image (b) DCT VQ compressed image at 42.728 dB.

Fig. 4. (a) Reconstructed image, category 1. (b) Packet loss map.
Fig. 5. (a) Reconstructed image, category II. (b) Packet loss map

Fig. 6. (a) Reconstructed image, category III. (b) Packet loss map

Fig. 7. (a) Reconstructed image, category IV. (b) Packet loss map