VECTOR QUANTIZATION OF IMAGES USING VISUAL MASKING FUNCTIONS *

Ramin Baseri
Center for Image Processing, ECSE Department
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

V. John Mathews
Department of Electrical Engineering
University of Utah
Salt Lake City, Utah 84112

ABSTRACT

This paper presents an image compression technique that incorporates visual masking functions in vector quantizer systems. Visual masking functions provide a description of the maximum amount of noise that can be present in an image, while remaining undetected when the image is viewed by an observer. The basic idea employed in this work is that of a spatially varying distortion measure which is defined to be zero where the error involved is below a threshold level defined by the visual masking function. A gradient based algorithm is used to generate the vector quantizer codebooks. Experimental results involving subband vector quantization and a perceptual masking function recently proposed by Safranek and Johnston are presented in this paper.

1 INTRODUCTION

A good data compression scheme should be capable of removing statistical and psychophysical redundancies present in the input waveforms. This paper presents a vector quantization (VQ) system that attempts to remove psychophysical redundancies present in images by making use of spatial masking functions. The key idea used in the development of this system is that of a space invariant distortion function that depends on a visual masking function. Even though the ideas discussed in this paper are applicable to any masking function, our experiments employed the masking function recently developed by Johnston and Safranek [8, 9]. This function is defined for a particular subband decomposition of the image and, consequently, the results presented in the paper involve subband vector quantization. Intraband vector quantization (forming the vector from within each subband) rather than interband vector quantization (forming the vectors using samples from different subbands) was used since it has been our experience that the former performs better than the latter in most situations. Allocation of the available bits among the various subbands was done using a systematic approach very similar to that recently proposed by Bradley, Stockham and Mathews [1]. Experimental results indicate that the use of perceptual masking function in a vector quantizer system reduces the perceived distortion in the coded image.

*This research was supported in part by NSF Grant No. MIP 9016331 and by a grant of computer time from the Utah Supercomputing Institute which is funded by the state of Utah and the IBM Corporation.

2 VQ EMPLOYING SPATIAL MASKING FUNCTIONS

Visual masking occurs when one visual stimulus affects the visibility of another. The response of the human visual system to stimuli is a function of the characteristics of the image in the vicinity of the stimuli. As a result, visual stimuli are masked ("hidden") by certain features in the background or in the vicinity of the stimuli [7, 6]. Several researchers have developed empirical functions that approximate the masking properties of the human visual system [6, 7, 8, 9]. Image coding systems employing such masking functions have resulted in coded images with very high visual quality.

Even though vector quantization is known to perform better than scalar quantization, no work has been done (at least to the authors' knowledge) that combines the concepts of vector quantization and visual masking. Our approach to developing a vector quantizer equipped with a masking function is to make use of a space variant distortion function which is based on the masking function. The distortion between an m-dimensional input vector $\mathbf{x}$ and another vector $\mathbf{y}$ is defined as,

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{m} e_i^2 \tag{1}$$

where

$$e_i = \begin{cases} 0 & ;\text{if } |\mathbf{x}_i - \mathbf{y}_i| < M_i \\ |\mathbf{x}_i - \mathbf{y}_i| - M_i & ;\text{otherwise} \end{cases}$$

$\mathbf{x}_i$ and $\mathbf{y}_i$ are the i-th elements of $\mathbf{x}$ and $\mathbf{y}$, respectively, M_i is the value of the masking function corresponding to the $i-th$ element of $\mathbf{x}$. Notice that, in effect, the distortion measure ignores differences between elements of the two vectors that are smaller than the threshold suggested by the masking function. One feature of such a definition of the distortion function is that it is a measure of how closely one vector resembles another vector when viewed by a human observer. This, of course, assumes that the masking function is a good indicator of the amount of distortion that the human eye can tolerate. Since the effectiveness of image compression schemes are determined by assessing the subjective quality of the coded images, it is only logical that a data compression system that attempts to minimize a distortion measure defined based on visual considerations will produce higher quality (as viewed by the observer) images as compared to schemes using distortion measures that are defined otherwise.

We will now present a gradient based algorithm for designing codebooks that are appropriate for use with our distortion function. The algorithm operates as follows. Starting with an initial codebook (denoted by $C(0)$), let $C(m)$ be the codebook obtained after encoding the m-th training vector. Also, let $C_i(m)$ denote the i-th code vector in $C(m)$. At the k-th iteration, find the codeword in $C(k-1)$ that is closest to $\mathbf{x}_k$, the k-th training vector. Let $C_L(k-1)$ be the closest codeword. Only $C_L(k-1)$ is updated at the k-th iteration. $C(k)$ is obtained by replacing $C_L(k-1)$ in $C(k-1)$ with

$$C_L(k) = C_L(k-1) - \frac{\mu}{2} \nabla_{C_L(k-1)} D(\mathbf{x}_k, C_L(k-1)). \quad (2)$$

Here μ is a positive constant in the range $0 < \mu \leq 1$ that controls the rate at which the code vector sequence converges to a steady state value. Also, $\nabla_{C_L(k-1)} D(\mathbf{x}_k, C_L(k-1))$ is the gradient of the distortion measure $D(\mathbf{x}_k, C_L(k-1))$ with respect to $C_L(k-1)$, and its i-th entry is given by

$$\nabla_{C_L(k-1),i} D(\mathbf{x}_k, C_L(k-1)) = \quad (3)$$
$$\begin{cases} 0 & ; \text{if } |\mathbf{x}_{k,i} - C_{L,i}(k-1)| \leq M_{k,i} \\ -2(|\mathbf{x}_{k,i} - C_{L,i}(k-1)| - M_{k,i}) \times \\ \qquad \text{sign}(\mathbf{x}_{k,i} - C_{L,i}(k-1)) & ; \text{otherwise.} \end{cases}$$

where $\mathbf{x}_{k,i}$ and $C_{L,i}(k-1)$ are the i-th entries of the $\mathbf{x}_k$ and $C_L(k-1)$, respectively and $M_{k,i}$ is the masking function associated with the i-th element of $\mathbf{x}_k$.

The above process is continued untill all the training vectors are exhausted or convergence is achieved in some sense.

3 SUBBAND CODING USING VISUAL MASKING FUNCTIONS

To demonstrate the usefulness of the ideas developed in the previous sections, we conducted experiments involving subband vector quantization and a slightly modified version of a perceptual threshold masking function recently developed by Safranek and Johnston [8, 9]. This function is defined on a sixteen band subband decomposition of the image. The masking function associated with the (x,y)-th pixel in the n-th subband as defined by Safranek and Johnston is given by

$$M(n, x, y) = Base(n) \times TexEnergy(x, y)^{0.034} \times BrightCorr(x, y) \quad (4)$$

In equation (4), $Base(n)$ is the root-mean-squared base noise sensitivity for subband n. The base sensitivity for each subband was empirically measured by Safranek and Johnston and can be found in [9]. Also, $BrightCorr(x, y)$ is a scaling factor which is dependent on the background intensity of the image and $TexEnergy$ is a measure of the nonuniform response of the human visual system to local activity in an image. This measure is defined as

$$TexEnergy(x, y) = \quad (5)$$
$$\sum_{n=2}^{16} MTFweight(n) \times Energy(n, x, y) + MTFweight(1) \times var((x, y), (x+1, y), (x, y+1), (x+1, y+1))$$

Where, $MTFweight(n)$ (Modulation Transfer Function weight) is the amplitude of the spatial frequency response of the human visual system at the center frequency of the n-th subband [2], $Energy(n, x, y)$ is the energy of the intensity of the target pixel and is defined as amplitude squared value of the (x,y)-th pixel in the n-th subband. $var(.,.,.,.)$, is evaluated by calculating the variance of the intensities over a $2 \times 2-pixel$ square block in the first subband and defined by the indices within the parentheses. For a full description of some of the conditions under which the empirical measurements of the parameters of the threshold model were made, see [8, 9]).

The only modification to the masking function in (4) that was made in this work was to set the brightness correction factor, $BrightCorr(x, y)$, to one uniformly. The modification was primarily motivated by the fact that all the images in our library are stored as density values (logarithm of the intensity values) and the nonlinear transformation does attempt to account for the differences in the response of the human visual system to different background intensities.

One of the difficulties involved in subband vector quantization is that of determining the optimum bit rate and vector size with which each subband must be coded such that the overall distortion in the reconstructed image is minimized. In our work, we have used a scheme recently developed in [1] for the optimal allocation of bits among the subbands and the determination of the vector size in each subband.

The key ideas employed in this scheme are as follows. It is assumed that the overall distortion of the image $D(\mathbf{r}, \mathbf{k})$ can be written as the sum of the distortions in the individual subbands, i.e.,

$$D(\mathbf{r}, \mathbf{k}) = \sum_{i=1}^{m} d_i(\mathbf{r}, \mathbf{k}) \quad (6)$$

where $d_i(\mathbf{r}, \mathbf{k})$ represents the distortion introduced in subband i as a result of vector quantization at rate $\mathbf{r}_i$ and vector size $\mathbf{k}_i$. $\mathbf{r}$ is an m-dimensional vector containing the rates corresponding to each subband signal and similarly, $\mathbf{k}$ is an m-dimansional vector containing the subband vector sizes.

Bradley [1] demonstrated that $d_i(\mathbf{r}, \mathbf{k})$ can be approximated by

$$d_i(\mathbf{r}, \mathbf{k}) = \beta_i(\mathbf{k})e^{-\gamma_i(\mathbf{k})r_i} \quad (7)$$

for image sequences when the Euclidean distortion measure is employed. Here $\beta_i(\mathbf{k})$ and $\gamma_i(\mathbf{k})$ are parameters that must be selected empirically. We have verified that the above approximation is reasonably accurate for our distortion measure as well.

The bit allocation problem can now be stated as that of selecting the bit rate $\mathbf{r}_i$ and vector sizes $\mathbf{k}_i$ for $i = 1, 2, \ldots, m$ such that

$$D(\mathbf{r}, \mathbf{k}) = \sum_{i=1}^{m} \beta_i(\mathbf{k})e^{-\gamma_i(\mathbf{k})r_i} \quad (8)$$

is minimized subject to the constraints that,

$$\sum_{i=1}^{m} \mathbf{r}_i = \text{Total bit rate} \quad (9)$$

and

$$\mathbf{r}_i \geq 0 \qquad ; i = 1, 2, \ldots, m \quad (10)$$

and a complexity constraint

$$\sum_{i}^{m} 2^{\mathbf{r}_i \mathbf{k}_i} \leq \text{Maximum codebook size.} \quad (11)$$

In order to further simplify the problem, we constrain the vector sizes to belong to a finite set. In our work we chose to select $\mathbf{k}_i$ to be from the set

$$\{(1 \times 1), (2 \times 2), (4 \times 4), (8 \times 8), (16 \times 16)\}. \quad (12)$$

The optimization process involves two sequential steps.

1. Optimization of bit rates using a projected gradient algorithm and the constraint set [1, 3].
2. Given the optimal bit rate, a pattern search algorithm [1] is used to find the the optimal vector sizes such that $D(\mathbf{r})$ is minimized.

Details of the optimization technique may be found in [1] and are not repeated here.

In this work, the first subband was coded using predictive vector quantization (including DPCM as a vector quantizer with vector size of 1 pixel). Higher subbands were coded using direct vector quantizers.

4 EXPERIMENTAL RESULTS

The usefulness of the technique presented in the paper was evaluated by performing several experiments with monochrome images which were quantized to 256 (0-255) shades of gray. The test image was chosen to be the "woman" image and was not a part of the training sequences. The test image is shown in Figure 1.

The test image was coded using a maximum codebook size of 512 code vectors at 0.5 bits/pixel using the masking function and is displayed in Figure 2. The same image was also coded without using any type of visual criteria using the same overall bit rate and maximum codebook size. Enlarged portions of the coded images obtained with and without the use of the masking function is displayed in Figure 3 and Figure 4. The image coded using the perceptual masking function appears to have better visual quality than the image coded without the masking function. Specifically it lacked a grid-like distortion which was present in the image not coded using the masking function.

Figure 1: The original test image.

5 CONCLUDING REMARKS

This paper presented an image compression technique that employs spatial masking functions. The technique makes use of two well studied topics in the fields of image coding and human psychophysics, namely vector quantization and spatial visual masking. An empirical masking function representing the noise threshold in image subbands was used to define a bound on the visually tolerated quantization noise resulting from subband vector quantization. Compressed images using this approach display definite improvements over the conventional subband vector quantization algorithms which make no use of the properties of the human visual system.

References

[1] J.N. Bradley, T.G. Stockham and V.J. Mathews, "Subband image coding with vector quantization," submitted to *IEEE Trans. Communications.*

[2] T.N. Cornsweet, *Visual Perception*, Academic Press, New York, 1970.

[3] P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, Academic Press, 1981.

[4] A.K. Jain, *Fundementals of Digital Image Processing*, Prentice Hall, 1989.

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

[6] J.O. Limb, "Distortion criteria of the human viewer," *IEEE Trans. on Systems, man, and cybernetics,* ' Vol. SMC-9, No. 12, pp. 778-793, December 1979.

[7] A.N. Netravali, B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE*, Vol. 65, No. 4, pp. 536-548, April 1977.

[8] R.J. Safranek, J.D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," *Proc. ICASSP 89, pp. 1945-1948, Glasgow, Scotland, May 1989.*

[9] R.J. Safranek, J.D. Johnston, and R.E. Rosenholtz, "A perceptually tuned sub-band image coder," *Proc. SPIE Symp. Human Vision and Electronic Imaging: Models, Methodes and Applications, Santa Clara, California, February 1990.*

[10] T.G. Stockham, "Image processing in the context of a visual model," Proc. IEEE, Vol. 60, No. 7, pp. 828-842, July 1972.

[11] Z. Xie, "Previsualized image vector quantization," *Ph.D. Dissertation* University of Utah, 1989.

Figure 2: Test image coded using the perceptual masking fuction.

Figure 3: An enlarged portion of the test image coded using the masking function.

Figure 4: An enlarged portion of the test image coded without using any perceptual criteria.