

IMPROVED DISTORTION MEASURES FOR IMAGE VECTOR QUANTIZATION

Anamitra Makur

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore, India

ABSTRACT: The role played by the distortion measure in a vector quantization image encoder is very important. In the following paper we suggest a general class of distortion function, the input-dependent weighted square error distortion, which is computationally simple and can be used to incorporate some psychovisual characteristics. Two techniques of using this distortion function have been discussed. Incorporating various classes depending on image activity in the distortion measure, better subjective quality can be achieved without added complexity, as our simulation studies have shown. This strategy of classified codebook is optimal and more general than the conventional sub-codebook method. We also show that incorporating dimensional emphasis, blockiness can be reduced in lieu of marginally more computation.

I. Introduction

Vector Quantization (VQ) has evolved as an efficient computation technique specially for low rate speech and image coding. From a pre-computed set of typical vectors/blocks called codebook, the VQ encoder finds the one best matching the input vector, and transmits the index of this best match. The VQ decoder simply looks up the codebook for the vector. Refer to [1] for the basics of VQ and to [2] for a review on VQ image encoder.

For a given training sequence and N , cardinality of the codebook, the LBG algorithm [3] designs the optimum vector quantizer, i.e., the quantizer which encodes the training sequence with minimum average distortion. A distortion function is used to measure this quantization distortion. If X be a K -dimensional input vector, and C , also K -dimensional, be a member of the codebook, then $d(X, C)$ is the distortion function (a scalar) associated with the VQ coder. The LBG algorithm works for any work done at the Dept. of EE, California Institute of Technology, Pasadena, with the financial support of Pacific Bell

such distortion function as long as $d(X, C)$ and the centroid of a set of vectors using this distortion function exists.

The centroid for a cluster or a set of input vectors is defined to be the vector (not necessarily unique) having a minimum average distortion between it and any other member of the set. Thus, for such a set B , the centroid C is given by

$$C = \min_Y \left[\begin{array}{c} E\{d(X, Y)\} \\ X \in B \end{array} \right]. \quad (1)$$

Each member of a codebook is a centroid, which is why the same symbol has been used to denote both.

II. Input-Dependent Weighted Square Error Distortion

The distortion measure almost universally used in image coding is the mean square error (MSE) distortion,

$$d(X, C) = \frac{1}{K} [X - C]^t \cdot [X - C], \quad (2)$$

where t indicates transpose, and vectors are column vectors. In case of MSE the distortion value is proportional to the square of the Euclidean distance between the two vectors, and the centroid happens to be the mean of the set.

Since the primary objective of an image coder is human viewing, the mathematical function used for computing the distortion is nothing but an attempted quantification of human visual discomfort towards quantization errors in the decoded picture. It has been recognized quite early that the MSE distortion is far from a true measure of the psychovisual distortion. This fact motivated the search for a computationally simple, yet more humanlike, distortion measure.

The MSE distortion is a special case of a class of distortion measures meeting the LBG requirement, namely the weighted mean square error (WMSE) distortion,

$$d(X, C) = \frac{1}{K} [X - C]^t \cdot W \cdot [X - C], \quad (3)$$

where W is the weight matrix of size $K \times K$. We suggest a general class of distortion measure by extending the WMSE func-

tion such that the weight matrix depends on the input vector. We call it the input-dependent weighted square error (IDWSE) distortion function, which is given by

$$d(X, C) = [X - C]^t \cdot W_X \cdot [X - C]. \quad (4)$$

The above function can be easily computed. In order for the centroid using the IDWSE distortion to exist too, we add a constraint that W_X be a diagonal matrix. Then the distortion in (4) becomes the sum of the distortions for individual dimensions with no cross-term, and it can be minimized by minimizing each dimensional component independently. If x_i and c_i are the i -th components of X and C respectively, and $W_X = \{w_{ii}(X)\}$, then the distortion for the i -th dimension, d_i , is

$$d_i = w_{ii}(X)(x_i - c_i)^2. \quad (5)$$

When this distortion is minimized, then we have

$$c_i = \frac{E\{w_{ii}(X)x_i\}}{E\{w_{ii}(X)\}}, \quad (6)$$

where the expectations are taken over the set B . Therefore, the centroid of a cluster is

$$C = \frac{E\{W_X \cdot X\}}{E\{W_X \cdot \mathbf{1}\}}, \quad (7)$$

where $\mathbf{1}$ is the unit column vector. Incorporating the above result in the LBG algorithm, it is then possible to design a VQ codebook optimum in the IDWSE sense.

III. Use of the Activity Classes

In order to improve the subjective quality, the image activity can be used to treat different parts of an image separately in an encoder. Some earlier image coders (not VQ) have used the activity index [4] or the level of activity [5], each of which needs a fair amount of computation. We propose using the IDWSE distortion function such that the weight matrix of equation (4) is of the form

$$W_X = a(X) \cdot I, \quad (8)$$

where $a(X)$ is a positive scalar function of the input vector, X , and is called the activity index of X . Negative or zero value of the activity index may lead to inconsistent distortion measure. For a spatial block of image, the activity index may be defined as the amount of detail, or changes in intensity, present in that block. In a typical image, these high-activity blocks are responsible for the quality of the image, and are usually more noticeable. Thus, a measure of the local activity gives us a chance to segment the image into different classes and to encode each

class differently.

The coder operates as follows. An input vector is assigned an activity index depending on its activity. The activity index should reflect the relative perceptual importance or visibility of the input vector. A codebook is then generated using this distortion function. It is possible to emphasize, or de-emphasize, any particular class of vectors in the codebook by simply raising, or lowering, the value of the activity index attached to this class. If n distinct classes are desirable, the activity index $a(X)$ shall take only one of n values. On the other hand, there could be infinitely many classes, such as a monotone continuous activity index.

The concept of careful encoding of the high-activity parts of an image has earlier been used in VQ in an empirical way. Briefly, the input vectors are classified into two (or more) classes, edge and shade, using an edge-classifier [6, 7, 8]. Separate sub-codebooks are designed for each class, and the final codebook is a concatenation of the sub-codebooks. The final codebook, however, is no longer optimal in the MSE or any other sense. Our scheme has two distinct advantages over the sub-codebook method. Firstly, the modified distortion measure, which needs marginally more computation, is used only during the codebook designing, which however is done off-line. The encoding process is exactly the same as for an MSE VQ coder, since $a(X)$ is constant while minimizing the distortion over the codebook. Secondly, the codebook in our case remains optimum for the IDWSE distortion measure, and the function $a(X)$ may be optimized independently using subjective criterion.

We propose a computationally much easier way to determine the activity index of a given vector than suggested in [4, 5]. On each image block, m pairs of randomly placed pixels are considered. The difference in the intensity value of each pair is added and the magnitude of this total difference is an indication of the image activity. Only $(2m - 1)$ additions are required to determine the activity of an input block this way. A scalar quantizer is then used to quantize this value so that an n -level activity index results.

IV Emphasis on the Block Boundaries

An annoying coding artifact in the low bit rate block image coding is blockiness, occurring in the block boundaries. For the MSE distortion each pixel of the input block is encoded with equal effort. However, in order to preserve the continuity of an image across the blocks, the peripheral pixels are needed

to be encoded more carefully. We propose use of the IUWSE distortion measure to achieve this effect. In this case the weight matrix becomes

$$W_X = \{d_{ii}\}, \quad (9)$$

where d_{ii} is independent of the input vector, X , and is called the emphasis factor for the i -th dimension. For the dimensions i corresponding to the interior pixels, $d_{ii} = 1$. The remaining diagonal elements corresponding to the pixels at the block boundaries are set equal to δ , where $\delta > 1$. The emphasis factor represents the relative importance of a dimension.

Because the weight matrix is independent of X , the centroid is simplified from equation (7) to be the arithmetic mean of the cluster. The distortion measure in the block boundary emphasis scheme increases both the codebook designing and the encoding computation. The number of elements not equal to 0 or 1 in W_X is $4\sqrt{K} - 4$. Therefore, an additional $4N(\sqrt{K} - 1)$ multiplications are required in encoding each input vector in this scheme. Taking δ to be integer power of 2, the additional multiplications could be avoided.

V. Simulation Results

Simulation results of the two IDWSE codebooks for low rate image coding have been compared with the performance of the MSE codebook, while keeping all other parameters the same. Firstly, the image activity has been computed. Square blocks of size 8×8 have been taken, and $m = 8$ pairs of random pixels are used. Figure 1 is the original image of Leena, possibly the face most experimented with. Figure 2 shows the image activity of the blocks of Leena (a darker block implies higher activity). Four activity classes have been defined by optimally segmenting the activity range, as discussed in [9, pp. 43-44]. The values of the activity index for the four classes have been taken as $a_i = 1 + (i - 1)\delta$ for $1 \leq i \leq 4$ for different positive values of δ . $\delta = 0$ is equivalent to the MSE distortion.

The simulation results show that with increasing value of δ , the edges and the contrasting blocks are indeed encoded better. Figure 3 shows average MSE/pixel of the decoded image from the IDWSE codebooks for various values of δ (bit rate 0.156 bit/pixel, $N = 1024$). Slightly smaller distortion than the MSE codebook case ($\delta = 0$) has been achieved for small values of δ . This is not a contradiction since the LBG algorithm promises minimum average distortion only for the training sequence. Though the MSE distortion is higher, improved subjective quality is observed for higher values of δ . Figure 4 shows the

decoded image from the MSE codebook (0.156 bit/pixel, SNR 26.88 dB). Figure 5 shows the decoded image from the IDWSE codebook with $\delta = 2$ (0.156 bit/pixel, SNR 26.81 dB). The edges are better represented in figure 5 than in figure 4.

The emphasis factor has also been used in the IDWSE distortion to reduce blockiness. As earlier, several codebooks have been designed for different values of the emphasis factor, δ . The MSE codebook is equivalent to $\delta = 1$. Figure 6 shows the average MSE/pixel of the decoded image for various values of δ . Though for small values of δ the IDWSE codebook shows a lower distortion than the conventional codebook, the subjective improvement becomes noticeable for higher values of δ . Figure 7 shows the decoded image using the IDWSE codebook with emphasis factor $\delta = 2.5$ (0.156 bit/pixel, SNR 26.81 dB). The blockiness of figure 7 is considerably less than that of figure 4. Moreover, the edges of figure 7 are better represented than figure 4.

The codebook generated by the IDWSE distortion incorporating emphasis factor has another advantage over the MSE codebook. It is desirable that all members of a codebook are utilized nearly equally often. To determine the utilization pattern of a codebook, we computed the variance of the number of times each member of the codebook is used for the training sequence. Figure 8 plots the variance versus the emphasis factor, δ , which shows that the variance of a VQ codebook reduces with an increase in emphasis factor.

VI. Conclusion

The IDWSE distortion, which is computationally simple and usable with the LBG algorithm, has been shown to offer decoded images of better subjective quality than the MSE distortion. With the activity index, better edge representation has been achieved for low bit rate VQ image coding without sacrificing anything. Further, introducing some positive emphasis at the block boundaries has produced a decoded image with less blockiness and better edge representation, while the encoder computation is marginally raised. Other modifications to incorporate human visuality in the IDWSE distortion can also be thought of. Moreover, it is easy to use the IDWSE codebooks with the improved VQ coders suggested by others in order to gain the advantages of both.

The author wishes to acknowledge the guidance he received while performing this work from Prof. Edward C Posner of Caltech



Figure 1: Original Image of Leena



Figure 4: Decoded Image from MSE Codebook



Figure 2: Image Activity of Leena



Figure 5: Decoded Image from Codebook with Activity Index

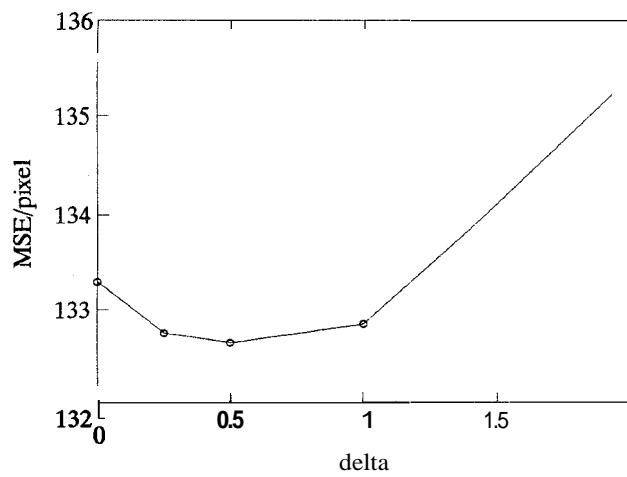


Figure 3: MSE Distortion for Codebooks with Activity Index

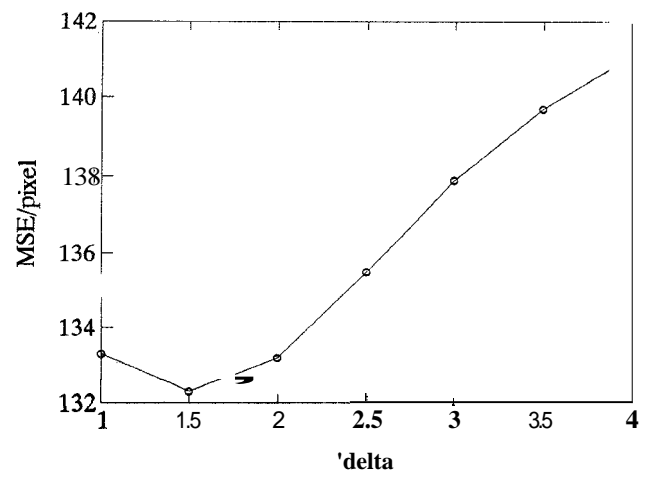


Figure 6: MSE Distortion for Codebooks with Block Boundary Emphasis



Figure 7: Decoded Image from Codebook with Block Boundary Emphasis

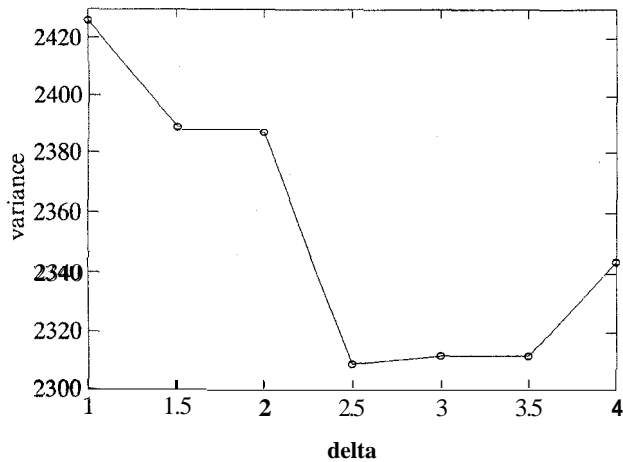


Figure 8: Utilization of Codebooks with Block Boundary Emphasis

References

- [1] R. M. Gray, "Vector Quantization," *IEEE ASSP Mag*, vol. 1, pp. 4-29, April 1984.
- [2] N. M. Nasrabadi, R. A. King, "Image Coding Using Vector Quantization: A Review," *IEEE Trans. Comm*, vol. 36, pp. 957-971, August 1988.
- [3] Y. Linde, A. Buzo, R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Comm*, vol. COM-28, pp. 84-95, January 1980.
- [4] J. I. Gimlett, "Use of 'Activity' Classes in Adaptive Transform Image Coding," *IEEE Trans. Comm*, vol. COM-23, pp. 785-786, July 1975.
- [5] W. Chen, C. H. Smith, "Adaptive Coding of Monochrome and Color Images," *IEEE Trans. Comm*, vol. COM-25, pp. 1285-1292, November 1977.
- [6] A. Gersho, B. Ramamurthi, "Image Coding Using Vector Quantization," in *IEEE Proc. Intl. Conf. ASSP*, 1982, pp. 428-431.
- [7] B. Ramamurthi, A. Gersho, "Image Vector Quantization with a Perceptually-based Cell Classifier," in *IEEE Proc. Intl. Conf. ASSP*, 1984, pp. 32.10.1-32.10.4.
- [8] B. Ramamurthi, A. Gersho, "Classified Vector Quantization of Images," *IEEE Trans. Comm*, vol. COM-34, pp. 1105-1115, November 1986.
- [9] A. Makur, "Low Rate Image Coding Using Vector Quantization," Ph.D. Thesis, California Inst. of Tech, 1990.