IMPROVEMENTS ON STOCHASTIC VECTOR QUANTIZATION OF IMAGES

L. Torres   **J. Salillas**

Dept. of Signal Theory and Communications

ETSIT - UPC

P.O. Box 30.002 - 08080 Barcelona, SPAIN

ABSTRACT

The vector quantization scheme has proven to be very effective in speech and image coding. One of the most important steps in the whole process is the design of the codebook. The codebook is generally designed using the LBG algorithm, that is in essence a clustering algorithm which uses a large training set of empirical data that is statistically representative of the image to be quantized. In order to avoid some inherent problems related to the generation of the codebook using the LBG algorithm, a stochastic generation of the codebook has already been proposed as an efficient alternative. In this case, the codebook is generated according to some previously defined model of the image to be encoded and not according to some specific data sequence. The objective of this paper is to report on new results using some improved models for the generation of the codebook and further refinements on the visual aspect of the encoded image. Good visual results are shown in the range of 0.5 - 0.8 bits/pixel. For completeness purposes, a summary of the original stochastic vector quantizer scheme is also provided.

1. INTRODUCTION

The vector quantization scheme has proven to be very effective in speech and image coding [1] [2]. In image vector quantization, the image is divided into small rectangular subimages typically of size 3x3 or 4x4 pixels. Each subimage is then compared with a set of standard subimages (codebook) and the codeword identifying the best match is transmitted. The codebook contains all the available codewords and the quantizer replaces the input subimage by the codeword that is more similar to the subimage to be coded. The euclidean distance is generally used as measure of similarity. Other distances are also used, although they will not be considered in this work. An optimal vector quantizer should ideally match the statistics of the input vector source [3]. However, if the statistics of an input image source are unknown, a training set representative of the expected input image source is used to design the vector quantizer. One of the most important steps in the whole process is the design of the codebook. The codebook is generally designed using the LBG algorithm [3] that is in essence a clustering algorithm which uses a large training set of empirical data that is statistically representative of the image to be quantized. This algorithm, although quite effective for practical applications is computationally very expensive and the resulting codebook has to be recalculated each time that the type of images to be encoded changes. Besides, the LBG algorithm only provides locally optimum codebooks that depend strongly on the images utilized to initialize the algorithm. On the other hand, for a codebook to be of practical value, it should perform well for a large class of images which were not part of the training data available for the codebook design.

In order to overcome the above explained problem concerning the design of the codebook, the stochastic generation of the codebook has been proposed [4]. The approach is to design the codebook according to some previously defined model for the image or type of images to be encoded. That is, to generate the training set according to the same model and not according to some specific data sequence as it is usually done. The stochastic generation of the codebook may provide two different advantages. First, the generation of the codebook is both conceptually and computationally very simple. Secondly, high compression rates can be obtained as a small codebook will be needed if the model has been well defined. If the image to be encoded follows closely the model used to stochastically generate the codebook, a very important reduction of the codebook size can be achieved, while retaining the vector dimension. That means that important reductions of the bit rate can be attained. On the other hand, the Stochastic Vector Quantization scheme can be made adaptive to follow the statistical changes of a particular subimage. Some similar approaches have already been used in speech coding [5]. The paper gives a short summary of the Stochastic Vector Quantization technique and presents new improvements for fixed-rate schemes in both visual quality and PSNR.

2. STOCHASTIC VECTOR QUANTIZATION

The only difference of the stochastic vector quantizer (SVQ) with respect to the conventional vector quantizer is the design of the codebook. In the SVQ the codewords are generated by stochastic techniques instead of being generated by a training set representative of the expected input image. In this context, the image data to be encoded, i.e., the subimages of size 3x3 or 4x4, are considered as realizations of a certain stochastic process. If the codewords are designed according to realizations of this process, the subimages to be encoded will find in the codebook, with a high degree of

probability, a very similar codeword, thus resulting in an efficient coding. The quality of the coding scheme will rely on the quality of the model selected to design the codebook.

The design of the codebook consists of the following stages: 1) Selection of a model, and 2) Synthesis of the codewords.

In order to generate the stochastic codewords, zero mean unity variance white-gaussian noise images of the same size as the subimages to be encoded, are passed through some shaping filter $H(z_1, z_2)$ whose output follows the selected model. Then, the filtered images are normalized to unity variance and used as codewords for the VQ scheme. Exactly as many stochastic images as the desired size of the codebook are filtered. With this approach, there is no need to use the LBG algorithm for the design of the codebook, thus avoiding the need of a very computationally expensive way of generating the codebook. The model may be selected before the coding process or the coding scheme must first generate the model from the observed signal to be coded. If the image to be encoded is not stationary (which is usually the case) and the model is selected from the observed signal, the scheme can be made adaptive by sending to the receiver the parameters of the shaping filter. This means that some overhead information has to be transmitted to the receiver. As the stochastically generated codebook allows a very important reduction of the codebook size, the overhead information does not represent a significant increase in the bit rate. It is worth mentioning that in this case, a different codebook has to be generated every time that a change in statistics of the input signal happens. However, the generation of the codebook through stochastic techniques is very fast and does not slow down the encoding/decoding process.

The model that is used in the design of the codebook has zero mean and unity variance. As the subimages to be encoded do not have zero mean, the first step in the encoding process is to remove the mean of each subimage. The means are subsequently coded using a 2D DPCM scheme. If after the removal of the means, the subimages are stochastically encoded with the codebook designed above, the variance information associated to each subimage will be lost. Thus, the assignment criterion between the subimage and the corresponding codeword has to be modified. The selected criteria, previously reported by Atal [6], is as follows. Given a certain input vector X (the subimage to be encoded), there exists for every codeword Y, a multiplicative factor γ that minimizes the distortion between the input vector and a new codeword γY. If a measure of distortion the euclidean distance between X and γY is defined

$$D = d(X, \gamma Y) = \| X - \gamma Y \|^2$$

then the scale factor γ can be found setting $\partial D / \partial \gamma = 0$ which gives

$$\gamma = \frac{X^T Y}{\| Y \|^2}$$

The corresponding multiplicative factor γ is coded with a 2D DPCM scheme and transmitted along with the index of the corresponding codeword.

3. IMPROVEMENTS ON SVQ

Although a variety of methods can be foreseen to improve the performance of the Stochastic Vector Quantization scheme (mainly by making the encoding adaptive), our objective here is to report on some advances in the non-adaptive scheme, in both visual quality and PSNR, using a different model than the one used in [4]. Quite good results for a bit rate of 0.8 bits/pixel were reported there using a very simply 2D AR image model. It is clear though, that a 2D AR model has the advantage of the simplicity, but on the other hand cannot account for the edge structure which appears in real-world imagery. If the AR model is estimated using natural images, for which the ratio between "monotone" and "edge" subimages is overwhelming, the model will be only able to produce slow and smooth transitions.

In order to improve the model, a much better previously reported stochastic model [7] has been successfully used in our encoding scheme. The codewords are generated according to the following procedure. The image is partitioned into a number of contiguous geometric regions, separated by a field of random lines. Elementary regions are assigned gray levels in such a way that the correlation coefficient of the gray levels within contiguous regions is equal to a given value ρ. The field of random lines is also adjusted so that the "edge density" (that is edges/pixel) is equal to a given value λ. This model is then approximated by means of spectral estimation techniques by a 2D ARMA model. Details can be found in [7]. The generation filter can then be shown to have the following form:

$$H(z_1, z_2) = \frac{1 + a_1(z_1^{-1} + z_2^{-1}) + a_2(z_1^{-1} z_2^{-1})}{1 - b_1(z_1^{-1} + z_2^{-1}) - b_2(z_1^{-1} z_2^{-1})}$$

Note that this model is estimated from artificial data, where the "edge" ratio can be controlled by the parameter λ, but it still retains the generation simplicity of the simple 2D AR model. The codebook is then constructed by filtering zero mean unity variance white noise images of the same size of the subimages to be VQ encoded with a filter defined by $H(z_1, z_2)$. Each of the filtered subimages is normalized to unity variance and then used as a codeword to encode the original image. Exactly as many stochastic images as the size of the codebook that wants to be designed, are filtered.

3.1 Mean and scale factor encoding

In a block coding system, it is extremely important to code accurately the means of the blocks. Small errors in the coding of adjacent blocks will produce a very visible blockiness effect. An errorless coding technique, called Interblock Noiseless Coding (IBNC) will be used for mean encoding. IBNC was originally proposed to encode search paths in a tree-search VQ in [8]. A bidimensional extension was used for mean-encoding in [9]. A brief explanation is given here for completeness purposes. Let us assume that strongly correlated means of D-bit length have to be transmitted. The means of adjacent subimages are likely to be very similar. In this case, the length L of the common prefix between the adjacent means and the (D-L-1) remaining bits can be transmitted. Note that the transmission of the (L+1)th bit is not necessary, as it can be assured that is different from the (L+1)th bit of the previously transmitted mean. The prefix lengths are encoded with a Huffman code, so that the rate is further reduced.

The scale factor γ is transmitted using a Huffman encoded 2D DPCM scheme, as small errors can be allowed for the encoding of the gains. The total bit allocation for the means and the scale factors for a subimage size of 5x5 is around 8 bits per subimage, since 6 bits are needed for the mean and between 1 and 2 for the scale factors. In order to have a security margin, 9 bits will be accounted for in the results that will be presented.

3.2 Post-filtering of the SVQ coded image

Although the quality of the SVQ is good for a fixed bit rate coder, a look at a zoomed SVQ image will reveal the effects of block coding, namely the "staircase effect" and the "grid noise". These effects can be largely removed with no further increase in the bit rate using the nonlinear space-variant filter proposed in [10]. A brief explanation of its operation follows.

After VQ decoding, the image is partitioned into square subimages (blocks) which are classified as: monotone, rough or edge, distinguishing between four edge orientations (0, 45, 90 and 135 degrees). If the block is monotone, it is left untouched. If it is rough, a 2D separable low-pass filter of length 3 is applied. Edge blocks are filtered with a one-dimensional low-pass filter of length 5 applied parallel to the edge. A PSNR improvement of about 0.5 dB for a bit rate of 0.8 bits/pixel is obtained using this filter, which does not fully reflect the quality of the visual improvement. The application of the filter is very fast and does not practically affect the encoding/decoding time.

4. RESULTS

We have performed a variety of experiments to show the goodness of the proposed coder. The well known 512x512 image Lenna has been selected as test image. The codeword size is of 5x5 and a total bit rate below 0.8 bits/pixel with a PSNR of 33.46 dB has been obtained. The same image encoded with a 8x8 fixed rate DCT scheme at the same rate gives a PSNR of 29.5 dB. Other test images show equivalent results. It is worth noticing the sharpness of the edges, taking into account that the system is not adaptive. The proposed scheme has been compared against VQ coding schemes with codebooks generated from actual image data and results show at least equal performance, and in some cases a better visual performance of the stochastic codebook, when compared to the codebook generated from actual data. Image 1 shows the original image Lenna and Image 2 shows zoomed Lenna coded at 0.8 bits/pixel before filtering. Image 3 shows zoomed Lenna after filtering and Image 4 shows the final decoded image at 0.8 bits/pixel after filtering.

5. CONCLUSIONS

A new non-adaptive fixed-rate vector quantizer encoding scheme has been presented and preliminary results have been shown. The design of the codebook has been based in a stochastic approach in order to match a previously defined model for the image to be encoded. Following this approach, the generation of the codebook is made extremely simply in computational load while yielding a very good visual quality for images coded in the range of 0.8 bits/pixel. Much better performance is expected for adaptive schemes.

References

[1] R.M.Gray, "Vector Quantization". IEEE ASSP Magazine, Vol. 1, April 1984.
[2] A. Gersho, R. Gray, "Vector Quantization and Signal Compression", Kluwer Academic Publishers, 1992.
[3] Y.Linde, A.Buzo and R.M.Gray, "An Algorithm for Vector Quantizer Design". IEEE Transactions on Communications, Vol. 28, January 1980.
[4] L.Torres, E. Arias, "Stochastic Vector Quantization of Images", International Conference on Acoustics Speech and Signal Processing, San Francisco, March 24-27, 1992.
[5] M.R. Schroeder, B. S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at very Low Bit Rates", Proceedings of ICASSP, Tampa, March 1985.
[6] B. Atal,"High-Quality Speech at Low Bit Rates: Multipulse and Stochastically Excited Linear Predictive Coders", ICASSP 86, Tokyo.
[7] J.W. Modestino, V. Bhaskaran, "Robust Two-Dimensional Tree Encoding of Images", IEEE Transactions on Communications, Vol. 29, N0.12, December 1981, pp. 1786-1798.
[8] D.L. Neuhoff, N. Moayeri, "Tree Searched Vector Quantization with Interblock Noiseless Coding", Proc. Conference Inf. Science Systems, March 1988.
[9] J. Salillas, N. Farvardin, "Adaptive Multi-Stage Vector Quantization for Image Compression", EUSIPCO 92, Vol. 3, Brussels, August 1992.
[10] B. Ramamurthi, A. Gersho, "Nonlinear Space-Variant Postprocessing of Block Coded Images", IEEE Transactions on ASSP, Vol. 34, October 1986.

Image 1. Original image Lenna.

Image 3. Zoomed Lenna at 0.8 bits/pixel after filtering

Image 2. Zoomed Lenna at 0.8 bits/pixel before filtering.

Image 4. Lenna coded with the stochastic codebook
at 0.8bits/pixel.