Minimum-Entropy Clustering and its Application to Lossless Image Coding

Farshid Golchin* and Kuldip K. Paliwal
School of Microelectronic Engineering
Griffith University
Brisbane, QLD 4111, Australia
F.Golchin, K.Paliwal@me.gu.edu.au

Abstract

The Minimum-Entropy Clustering (MEC) algorithm proposed in this paper provides an optimal method for addressing the non-stationarity of a source with respect to entropy coding. This algorithm clusters a set of vectors (where each vector consists of a fixed number of contiguous samples from a discrete source) using a minimum entropy criterion. In a manner similar to Classified Vector Quantization (CVQ), a given vector is first classified into the class which leads to the lowest entropy and then its samples are coded by the entropy coder designed for that particular class. In this paper the MEC algorithm is used in the design of a lossless, predictive image coder. The MEC-based coder is found to sigificantly outperform the single entropy coder as well as the other popular lossless coders reported in the literature.

1 Introduction

The past two decades have witnessed significant advances in the field of image coding. There are currently a wide variety of coding solutions available for different image coding requirements. The majority of these advances have been made in the area of lossy image coding where the distortion introduced by the coding process is tolerable within limits. However in many cases, such as some medical imaging or astronomy, no alterations whatsoever to the original image can be tolerated. This type of coding is referred to as lossless or reversible coding.

Lossless coding is utilized in cases where images are likely to undergo further processing once decompressed, or where the coding artefacts may produce potentially hazardous results (e.g., medical diagnosis). The main drawback of lossless coding is the reduced compression ratios it produces when compared to lossy coding. There has been some work in recent years towards developing lossless coders with better compression ratios.

There are two main types of approaches to lossless image coding. The first and the more conventional approach is that of predictive coding. The operation of the coder can be divided into two main sections. First the redundancy in the image is removed using a predictor and then the residual signal is encoded by an entropy coder. The other significant type of approach to lossless coding is the multiresoloution approach which has become more prominent recently [2],[3],[4]. Although both types of approaches rely on entropy coding and can benefit from the Minimum-Entropy Clustering (MEC) algorithm described in this paper, we will focus our discussions on the predicitive coding scenario.

*The first author is supported in part by a CSIRO division of Telecom. and Indust. Physics Postgraduate Scholarship

Adaptivity of the predictor and entropy coder has proven to be the key to better performance in lossless coding. Adaptive coders are better equipped to handle the non-stationarities encountered in natural images. In the simple example of Figure 1, adaptivity can be incorporated into the predictor and/or the entropy coder, and will almost certainly improve the performance of the coder.

So far, many adaptive lossless coders have relied on the use of 'contexts' [1] or other similar concepts. A context defines the particular characteristics of a small neighborhood of pixels. The characteristics may be defined in the original picture or in DPCM domain. Most often, contexts are used to indicate the level of activity in a particular region. This indication of the level of activity is then used to select an appropriate entropy coder for that particular region. Thus the entropy coding is made adaptive.

The use of contexts has produced quite promising lossless coding results. However, these contexts are generally emipirically defined and not optimized in any fashion. The algorithm proposed in this paper allows us to define contexts in an optimal fashion which minimizes the bit-rate of the coder. We propose a clustering algorithm to design a classifier which in turn selects an appropriate entropy coder for each region of the image. This algorithm will be referred to as Minimum-Entropy Clustering (MEC). The applications of the MEC algorithm are by no means limited to predictive, lossless coding; however, the simple example shown in this paper clearly demonstrates its potentials.

2 The Minimum-Entropy Clustering Algorithm

Our aim in the use of the Minimum-Entropy Clustering (MEC) algorithm is to classify blocks of samples and then encode the samples in each block using a suitably designed entropy coder. This classification and the design of the entropy coders should be performed in a way such that the overall entropy is minimized. In this paper we propose an iterative clustering algorithm which achieves these goals in a way that can be compared to 'codebook design' for a Vector Quantizer (VQ) [6].

In order to design a coding system with N entropy coders, the MEC algorithm operates as follows:

1. *Initialization*: N probability distribution functions (PDF's) are defined. These PDF's will define the intial classes.
2. *Minimum-Code-Length classification*: Each block of samples $B = (b_0, b_1, ..., b_{m-1})$ is classified as belonging to class C such that the code length (after entropy coding) $L = -\sum_{i=0}^{m-1} log_2\, p(b_i|C)$ is minimized. This is similar to a nearest neighbour selection in VQ design.
3. *Re-Estimate class statistics*: Estimate the new class PDF's. This will ensure that the class statistics are matched to those of the samples in that class and hence the entropy is further reduced. This step is similar to a centroid calculation in VQ design.
4. *Iteration*: Stop if a maxmimum number of iterations is reached or the classes have converged. Otherwise, go to step 2.

Steps 2 and 3 form the core of the MEC algorithm. In each of these two steps, the overall entropy is reduced. Although there is no guarantee of convergence to a global minimum, it was found that the algorithm quickly converges to a satisfactory minimum. The speed of the convergence depends greatly on the initialization procedure used. However, in almost all cases, it was found that the algorithm converges to the same minimum regardless of the initial settings.

There are a number of choices available for the initial classification. However, a more suitable definition of the initial classes can significantly reduce the number of iterations required. Since we wish to group together blocks with similar statistics, a reasonable choice for initial classification would be classification based on the variance of the blocks. To do so, we choose a classification similar to that used by Chen and Smith [5]. The variance of each block is estimated and the blocks are listed in the order of increasing (or decreasing) variance.

Using the sorted list, $m-1$ threshold values (of variance) are selected and used to classify the blocks into m classes. After this classification, the class PDF's are estimated and used to define the initial classes.

This algorithm may be used in either a parametric or a non-parametric form. In its parametric form, all distributions are modeled as Generalized Gaussian (GG) distributions [7] whose shape parameter and variance are estimated. In this case, step 3 of the algorithm becomes a maximum-likelihood estimation of the shape parameter and variance. The main disadvatage of using the parametric form of the MEC algorithm is the additional processing required in the calculation of probabilities (Step 2) and the parameter estimation (Step 3).

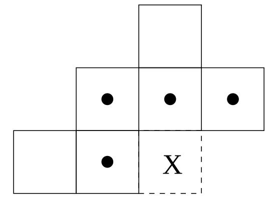

Figure 1: Causal Neighborhood of pixel X - The circle marks pixels used in prediction)

In the non-parametric version of the MEC algorithm, frequency tables are maintained for each of the classes. These frequency tables are updated in step 3 of the algorithm. After the MEC algorithm has converged, these frequency tables are used to design the entropy coders. It is for this reason that the frequency tables must also be known at the decoder. In the parametric case, the probabilities can be efficiently described to the decoder by transmitting two parameters (shape and variance) per class.

In the non-parametric version, the frequency tables must be explicitly transmitted. To reduce the amount of information, we may take adavantage of the symmetry in the frequency tables. One half of each frequancy table is quantized using 6-8 bits and then transmitted to the decoder. If adaptive entropy coders are used, the probability tables are quickly adapted to match the source statistics and any mismatches caused by quantization rapidly disappear.

There is a trade-off between the parametric and the non-parametric versions of the MEC algorithm. The non-parametric version offers faster execution times at the expense of added side-information (frequency tables). However, it should be noted that in terms of compression the two versions of the MEC algorithm produce quite similar results. The results in this paper have been obtained using the non-parametric version of the MEC algorithm.

3 An MEC-based Lossless Image Coder

As mentioned previously, adaptivity has proven to be a key reuirement in achieving high compression ratios in both lossy and lossless coding. Different areas within images contain different levels of activity. Smooth areas within an image contain less activity whereas strong textures and edges have higher activity levels. In the DPCM domain, these levels of activity are reflected in the values of local variance.

In this paper we propose the use of the MEC algorithm to provide adaptivity in the entropy coding stage of a lossless image coder. Blocks of the DPCM samples are classified and then encoded using an entropy coder designed to match the statistics of the particular class. The classification decision is based on the same Minimum-Code-Length criterion used in Step 2 of the MEC algorithm.

The decoder must also be made aware of the classification information. Hence, a classification table is transmitted as side information. This side information will require at most $log_2 N$ bits per block, where N is the number of classes. However, since all classes do not have the same probability of occurence, the classification table is also entropy coded to reduce the side information.

The predictor used in the DPCM coder is a fixed, 4th-order optimum linear predictor. The pixels used for prediction are the four nearest pixels which lie in the causal (previously transmitted) neighbourhood of the pixel to be encoded. This neighborhood is illustrated in Fig. 1.

The predictor has been designed to minimize the prediction error in a Mean-Squared-Error (MSE) sense. Although the optimal choice of a predictor in this case would be a predictor which minimizes entropy rather than MSE, the design of such a predictor is a time-consuming task which usually provides little additional gain. Especially since we are using more than a single entropy coder, the process of finding a minimum entropy predictor is made even more difficult. It is for this reason that we have chosen the MSE-optimized predictor (the prediction is rounded off to the nearest integer) which can be found by solving the familiar Wiener-Hopf equations.

In another paper [8], we examine how adaptive prediction may be used to achive even better compression results. The main aim of this paper is to demonstrate the advantages of MEC-based coding over a single entropy coder.

4 Coding Results

The coder described in the previous section was tested on a number of different monochrome images. The coding results are listed in Table 1. The bit-rates quoted in the table are based on actual file sizes and include overheads such as the classification tables and frequency tables (used in the non-parametric MEC).

Experiments were performed using different numbers of classes (entropy coders) and different block sizes. Using 16-classes and block sizes of 8x8 samples was found to produce the best results. If the blocks are made smaller or the number of classes are increased, then the overheads can become prohibitively large. On the other hand if the block sizes are made too large or the number of classes is reduced, the coder becomes less adaptive and performance begins to deteriorate.

Using 16 classes and 8x8 blocks, the classification tables contributed 0.06 bpp to the overall bit-rate. The frequency tables were responsible for 0.01 bpp of the total bit-rate for the 512x512 images and around 0.04 bpp for the 256x256 images.

Figure 2: Test Images (As listed in Table 1, starting from left to right, top to bottom)

The coder named Single-EC is a DPCM coder using the same predictor as the MEC-based coder and only a single entropy coder. CREW [2] is a transform-based (wavelet) lossless image coder developed at the RICOH California Research Centre. Since the CREW coder was not available for testing, the CREW results reported in Table 1 are limited to those reported by Zandi et al. in [2].

A set of 16 test images was used for testing the MEC-based coder. Thumbnail-sized replicas of the test images can be found in Fig. 2. The first 12 images are a collection of 512x512 pixel, 8-bit greyscale images, commonly used for testing image codecs. The last four images are 256x256 pixel, 8-bit greyscale Magnetic Resonance Images (MRI). Medical imaging is perhaps the largest single field of application for lossless image coding and hence it is only logical to include such medical images in testing lossless image coders.

The MEC-based coder was found to outperform the other three coder used in this comparison by minimum of between 0.02 bpp and 0.43 bpp depending on the test image. As expected, the highest gains were realized for the images which contained the largest variations within them (eg. barbara, crowd and vegas). In the four MR images, significant gains were observed for the mr2,mr3 and mr4 images, while the mr1 image displayed minor coding gains. This can attributed to the low level of detail within the mr1 image.

Considering the simple non-adaptive linear predictor being used the coding results of the MEC-based coder are very encouraging. We should also note that all of the coders quoted in Table 1, outperform the lossless JPEG standard [11], the results for which are listed in the last column of Table 1.

Coded	Coding Method			
Image	MEC	Single-EC	CREW [2]	JPEG
barbara	4.91	5.27	-	5.53
couple	4.78	4.92	4.91	5.17
crowd	4.02	4.36	4.26	4.69
lax	5.75	5.89	5.97	5.95
lena	4.21	4.44	4.34	4.64
man	4.54	4.80	4.73	4.93
fruit	4.66	4.79	-	4.91
hat	3.94	4.15	-	4.46
vegas	4.16	4.60	-	4.81
woman1	4.69	4.96	4.82	5.04
woman2	3.21	3.53	3.38	3.68
boat	4.44	4.69	-	4.89
goldhill	4.75	4.91	-	5.05
zelda	3.99	4.11	-	4.29
mr1	3.11	3.13	-	3.16
mr2	5.56	5.79	-	5.86
mr3	5.31	5.50	-	5.57
mr4	5.24	5.41	-	5.46

Table 1: A comparison of lossless compression results (quoted in bpp)

5 Conclusion

We have proposed a clustering based scheme for the optimal encoding of a decorrelated DPCM image using multiple entropy coders (MEC). The MEC algorithm provides an optimal method of using mutltiple entropy coders. Its advantages are clearly demonstrated by the coding results obtained using a simple DPCM coder. Its applications are however, by no means limited to lossless DPCM coding. The MEC algorithm may be used in combination with other lossless transforms such as the S+P transform [3] or the RTS transform used in CREW [2].

In another paper [9], we have investigated the use the MEC algorithm with the novel prediction scheme proposed by Wu and Memmon [12] in their coder named CALIC. The use of multiple entropy coders was found to improve the state-of-the-art performance of CALIC to higher compression ratios. The use of multiple entropy coders has also been investigated for lossy subband image coding [10]. These areas provide a large scope for future reasearch into the applications of this simple algorithm.

References

[1] T.V. Ramabadran and K. Chen, "The use of contextual information in the reversible compression of medical images," *IEEE Trans. Medical Imaging*, vol. 11, pp. 185-195, June 1992.

[2] A. Zandi, J.D. Allen, E.L. Schwartz and M. Boliek, "CREW: Compression with reversible embedded wavelets," *Proc. IEEE Data Compression Conference*, Snowbird, Utah, March 1995.

[3] A. Said and W.A. Pearlman, "Reversible image compression via multiresolution representation and predictive coding," *Proc. SPIE VCIP Symposium*, Cambridge, MA, 1993.

[4] A. Said and W.A. Pearlman, "An image multiresolution representation for lossless and lossy compression," *IEEE Trans. Image Processing*, vol. 6, pp. 243-250, June 1996.

[5] W.H. Chen and C.H. Smith, "Adaptive coding of monochrome and color images," *IEEE Trans. on Commun.*, vol. COM-25, pp. 1285-1292, November 1977.

[6] Y. Linde, A. Buzo and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Comm.*, COM-28, pp. 84-95, January 1980.

[7] N. Farvardin and J.W. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless sources," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 485-497, May 1984.

[8] F. Golchin and K.K. Paliwal, "Clustering-based adaptive prediction and entropy coding for use in lossless image coding," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, October 1997.

[9] F. Golchin and K.K. Paliwal, "Context Classification and Adaptive Prediction for Lossless Image Coding," *Submitted to IEEE Region 10 Conference - TENCON97*, To be held in Brisbane, Australia, December 1997.

[10] J.M. Lervik and T.A. Ramstad, "Optimality of multiple entropy coder systems for non-stationary sources modelled by a mixture distribution", *Proc. of the IEEE 1996 International Conference on Acoustics, Speech and Signal Processing*, Atlanta, Georgia, pp. 1874-1977.

[11] G.K. Wallace, "The JPEG still picture compression standard," *Comm. ACM*, vol. 34, pp. 30-44, April 1991.

[12] X. Wu and N. Memmon, "CALIC - A context based lossless image codec," *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, Atlanta, GA, pp. 1891-1894, May 1996.