Classified Adaptive Prediction and Entropy Coding for Lossless Coding of Images

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

Natural images often consist of many distinct regions with individual characteristics. Adaptive image coders exploit this feature of natural images to obtain better compression results. In this paper, we propose a classification-based scheme for both adaptive prediction and entropy coding in a lossless image coder. In the proposed coder, blocks of image samples (in PCM domain) are classified to select an appropriate linear predictor from finite set of predictors. Once the predictors have been determined, the image is DPCM coded. A second classification is then performed to select a suitable entropy coder for each block of DPCM samples. These classification schemes are designed using two separate clustering procedures which attempt to minimize the bit-rate of the encoded image. The coder was tested on a set of monochrome images and was found to produce very promising results.

1 Introduction

Image coding is an essential component of many digital transmission and storage systems. Generally speaking, image coding algorithms can be divided into two categories: lossy and lossless. Lossy compression techniques which are by far more popular, can result in minor changes (coding artifacts) which are not tolerable in certain applications. In areas such as medical image coding or some satellite imaging, coding artifacts can have potentially dangerous consequences or can result in the loss of data obtained at a great cost.

It is in these areas where lossless image coding methods are utilized. Although lossless methods offer significantly smaller compression ratios, the fact that they preserve the original data has made them essential. In recent years, a great deal of the research in image coding has been concentrated on lossy methods. The development of better transforms, quantizers and particularly adaptivity in coders has resulted in significant advances in this area.

The motivation for this work is drawn largely from the success that adaptive coders have demonstrated in lossy image coding. An adaptive coder is able to adapt its operation to the local characteristics within the different regions of an image. Adaptive lossless image coding is by no means a new area. The use of "contexts" [5] in various forms has demonstrated significant improvements in lossless compression results. However, many of these techniques are based on heuristics and often do not exploit the full advantages of adaptivity.

In its simplest form, a lossless DPCM-based coder is implemented using a fixed predictor and a fixed entropy coder. Adaptivity can be provided by making either one of these components or both of them adaptive. An adaptive predictor is able to better adapt to the local characteristics of various textures or the direction of edges and hence produce more accurate predictions. An adaptive entropy coder, on the other hand, can adjust its characteristics to better match the local statistics of an image and produce shorter-length codes.

The coder proposed in this paper provides both adaptive prediction and adaptive entropy coding. Both types of adaptivity are implemented through classification. The coder has a finite set of predictors and entropy coders available to it. For each small region (block) in the image, a decision is made as to which predictor and which entropy coder should be used. In the following sections we outline the methods used for designing the predictors, entropy coders and their respective classification schemes.

2 Classified Adaptive Prediction

As mentioned in the previous section, we aim to produce a classification-based coder, in which blocks of images are classified to select an appropriate predictor and entropy coder for each block. In this section, we describe how the predictors are designed.

The predictors used are 4-th order linear predictors. In these predictors, the prediction of the current pixel value is a linear combination of the values of the neighboring 4 pixels. The problem lies in determining the prediction coefficients and then adapting these coefficients to the local characteristics of the various regions in images. We should also note that the pixels used for prediction must lie in the causal neighborhood of the pixel to be predicted (ie.

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

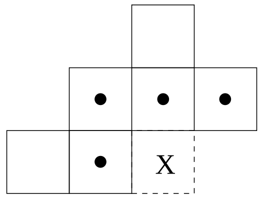

Figure 1: Causal Neighborhood of pixel X - The circle marks pixels used in prediction)

transmitted before the current pixel). The pixels used for prediction are marked in Fig. 1.

The aim of lossless coding is to encode an image losslessly and using the minimum number of bits. Hence, the objective in predictor design should be to reduce the entropy of the prediction residual. However, designing minimum-entropy predictors is a demanding task which has only demonstrated marginal gains over mean-square-error (MSE) optimized predictors. The MSE-optimized linear predictors can be easily designed by solving the Wiener-Hopf equations. It is for this reason that the MSE-optimized predictors were chosen for use in this paper. But if required, these predictors can be easily replaced by their minimum-entropy counterparts.

We use a clustering algorithm, in which the classification scheme and the predictors are designed in an iterative fashion. In this fashion, blocks of images are classified into a finite number of classes (typically 4), where each class is allocated its own predictor. This is done in a way such that the overall prediction error and hopefully the entropy are minimized. This procedure is similar to that used in Vector Quantizer (VQ) design [4].

The clustering algorithm operates as follows:

1. *Initialization*: Assign each class a set of predictor coefficients. A good choice for initial predictor coefficients is that of a single coefficient set to 1.0 and the others set to 0.0. This is done so that each predictor reflects a strong correlation in a particular direction (e.g., vertical, horizontal and two diagonals).
2. *Minimum-Prediction-Error classification*: Each block of samples $B = (b_0, b_1, ..., b_{m-1})$ is classified as belonging to class P such that the prediction error is minimized. This is similar to a nearest neighbor selection in VQ design.
3. *Calculate predictor coefficients*: For each class, a new optimum linear predictor is calculated. This step is similar to a centroid calculation in VQ design.
4. *Iteration*: Stop if a maximum number of iterations is reached. Otherwise, go to step 2.

In the above algorithm, each of the Steps 2 and 3 reduce the overall prediction error in a MSE sense. This reduction is also reflected in the entropy of the residue. Although there is no guarantee of convergence to a global minimum, it was found that 5-6 iterations were sufficient for the algorithm to approach a suitable local minimum. Jafarkhani et al. [3] have used a similar procedure for designing a spectral classification scheme for use in lossy wavelet-based coders.

At this stage, the image blocks can be classified and appropriate predictor coefficients have been calculated and are available for each class. In the next section we briefly describe how the adaptive entropy coders designed through the use of another clustering algorithm.

3 Classification-Based Adaptive Entropy Coding

In the previous section, we described how to design the predictors which are used in the lossless image coder. Through classification, the coder is able to select an appropriate predictor which best suits the directionality of the textures and edges within images.

Once the predicted values are removed from the original pixel values, we are left with the residual image which needs to be entropy coded. However, the residual image, which is largely decorrelated, contains regions with vastly different statistics. Smoother regions in images display low activity levels (low local variance); while, near edges and strong textures, the residual samples have much higher activity levels.

In another paper [2], we propose an algorithm which performs adaptive entropy coding of source by classifying blocks of samples into a finite set of classes and then encoding the members of each class by an appropriate entropy coder. The classifier design is quite similar to that described in the previous section. However, Steps 2 and 3 have been replaced by appropriate substitutes which aim to minimize the overall entropy and hence the bit-rate.

The algorithm was named Minimum-Entropy Clustering (MEC) and can be described in the following steps (for a coding system with N entropy coders):

1. *Initialization*: N probability distribution functions (PDF's) are defined. These PDF's characterize the initial 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 neighbor selection in VQ design.
3. *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 maximum 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. The speed of the convergence of this algorithm depends greatly on the initialization procedure used. For the purpose of this paper, an initialization scheme similar to the subband classiffication devised by Chen and Smith [1] was utlitized. In this initialization scheme, the variances of all blocks are calcualted and sorted in ascending order. After sorting, $N-1$ thresholds are determined such that the variance values are classified into N equally populated classes. Then, each image block is classified according to its variance and the PDF's of the N classes are estimated and used in Step 1 of the above algorithm. For a more thorough description of the MEC algorithm, the reader is referred to [2].

4 Coding Results

A lossless image coder was designed using the classified prediction and entropy coding methods described in the previous section. The coder was tested on a number of 8-bit monochrome pictures containing 256 grey levels. The first 12 images are commonly used coding test images which are 512x512 pixels in size. In addition, medical images were used for testing the coder. For this, Magnetic Resonance Images (MRI) of patients' heads were used. The MRI images are 256x256 pixels in size. Fig. 2 contains thumbnails of all the images used in testing.

It was found that the best results were obtained using 4 predictor classes and 16 entropy coders. The classification was performed on blocks of 8x8 pixels. In this coding scheme, the decoder must also be made aware of the classification decisions and hence a classification table must be transmitted to the decoder. For each block of 8x8 pixels, the classification table contains two entries, one entry specifies the predictor used for that block while the other entry describes the entropy coder used.

Using the above-mentioned number of classes and block size, the classification table contributes a maximum of 0.1 bpp to the overall bit-rate. Another set of overheads, results from transmitting the predictor coefficients for each class and the descriptions of PDF's of the entropy classes as described in [2]. This second overhead contributes a relatively small amount of approximately 0.015 bpp to the overall bit-rate (for 512x512 images). The overheads are written to a separate file and taken into account when calculating the overall bit-rate.

The compression results using the classified adaptive prediction and entropy coding (CAPEC) scheme are listed in Table 1. The compression results from three other lossless coders are also made available for comparison. The second column of the table marked Fixed-PE contains the results obtained using a single, fixed linear predictor and entropy coder (designed for each image). This is equivalent to having only 1 predictor class and 1 entropy coder.

The third column of Table 1, contains the coding results quoted for CREW [7]. CREW is a transform-based (wavelet) lossless image coder developed at the RICOH California Research Center. This coder was chosen for its good performance as well as the availability of its coding results for a similar set of images. Unfortunately, the CREW coder was not available for testing and hence the listed results are confined to those reported by the authors in [7]. The last column of Table 1 lists results obtained using the JPEG standard [6] coder in its lossless mode.

From the results in Table 1, it is clear that the coder proposed in this paper significantly outperforms the other three coders. The advantage of the Fixed-PE coder over the JPEG coder is due to its better predictor. Even though both coders use a fixed linear predictor, the predictor in the Fixed-PE coder is designed for the image being encoded and not unexpectedly outperforms the JPEG predictor.

CREW demonstrates a significant improvement over both the JPEG and Fixed-PE coders. This can be largely attributed to the adaptive entropy coding used in CREW. However, the proposed CAPEC system clearly outperformed CREW in terms of compression. To be fair to the CREW system, we should note that it offers other advantages in being a lossy and lossless coder in one. CREW's excellent lossless coding prefomance notwithstanding, it has not been optimized solely for lossless coding.

The CAPEC system significantly outperforms all other coders used in this test. This superior performance is due to the adaptivity offered by this coder. Particularly for images such as 'barbara', which contain strong and highly directional textures, the adaptive prediction greatly improves the coding performance. The coding result of 4.43 bpp for the barbara image is the best the authors have encountered in the literature. The coding results for other test images are very competitive with other state-of-the-art lossless coders found in the literature.

Among the tested coders, the CAPEC system was also the most computationally intensive system. This is the price we pay for adaptive prediction and entropy coding. However, in cases of encoding a particular type of image (eg. X-ray, MRI, etc.), a set of predictors and entropy coders can be trained "off-line" using a training set of images and hence eliminate the "on-line" time consumed by the optimization procedures.

5 Conclusion

We have proposed a DPCM image coder which is adaptive in both prediction and entropy coding. The adptivity is provided through classification. Blocks of image pixels are classified and consequently encoded using a particular predictor and entropy coder.

Two clustering algorithms for the design of the predictors and the entropy coders were proposed. The coder was tested on a set of test images which included 4 MR images. The test results clearly demonstrate the advatages of adaptive prediction and entropy coding. The proposed system clearly outperformed other coders used in the test.

The only significant drawback of the classified adaptive prediction and entropy coding scheme proposed in this paper is its computationally intensive training procedure (similar to codebook design in VQ terminology). However, as usually done in VQ, the predictors and entropy coders may be trained "off-line". Since most often lossless coders are only used to encode particular types of images such as

Coded Image	Coding Method			
	CAPEC	Fixed-PE	CREW [7]	JPEG
barbara	4.43	5.27	-	5.53
couple	4.59	4.92	4.91	5.17
crowd	3.92	4.36	4.26	4.69
lax	5.65	5.89	5.97	5.95
lena	4.06	4.44	4.34	4.64
man	4.41	4.80	4.73	4.93
fruit	4.43	4.79	-	4.91
hat	3.84	4.15	-	4.46
vegas	4.10	4.60	-	4.81
woman1	4.54	4.96	4.82	5.04
woman2	3.14	3.53	3.38	3.68
boat	4.23	4.69	-	4.89
goldhill	4.65	4.91	-	5.05
zelda	3.86	4.11	-	4.29
mr1	3.06	3.13	-	3.16
mr2	5.44	5.79	-	5.86
mr3	5.18	5.50	-	5.57
mr4	5.10	5.41	-	5.46

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

chest X-Rays or CT-scans, the use of coders trained "off-line" for a similar set of images will only result in a minor degradation in coding performance in return for a much smaller computational burden.

We should also note that the computational load of this system is not balanced equally between the coder and the decoder. A considerably larger portion of the computational burden lies with the encoder while the decoding can be performed in relatively few operations.

Future research will concentrate on the "off-line" training of predictors and entropy coders. We will also examine other (possibly faster) methods of optimizing the predictors and entropy coders.

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

6 Acknowledgements

The authors would like to thank the Image Processing Group at UMDS, Guy's Hospital in the United Kingdom for supplying the MR images.

References

[1] 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.

[2] F. Golchin and K.K. Paliwal, "Minimun-entropy clustering and its applcation to lossless image coding," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, October 1997.

[3] H. Jafarkhani, N. Farvardin and C.C. Lee, "Optimal two-dimensional spectral classification and its application in wavelet coding of images," Proc. Conference on Information Sciences and Systems (CISS-95), March 1995.

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

[5] 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.

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

[7] 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.