

# A CONTEXT-BASED ADAPTIVE PREDICTOR FOR USE IN 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

In this paper, we propose a context-based adaptive predictor for use in lossless image coding. Most often, lossless image coders utilize non-adaptive linear predictors for the sake of simplicity and to reduce the complexity of the coder. In DPCM-based lossless image coders, adaptivity can result in significant improvements in the performance. However, adaptive prediction is faced with a number of problems chiefly its extensive computational demands. The predictor proposed in this paper allows for a lower computational cost while guaranteeing the stability of the predictor. The Context-Based Adaptive Predictor (CBAP) was found to outperform or at least perform equally as well as the optimum linear predictor for a variety of test images. We should also note that designing an optimum linear predictor requires some knowledge of the image prior to coding while the CBAP requires no such knowledge and operates "on-the-fly".

## 1. INTRODUCTION

Lossless coding is an area of image coding with many important applications. In cases such as medical imaging, although compression is desirable, the usual side effects of lossy compression can not be tolerated. It is in such cases that lossless image coders are used.

There are existing standards and pseudo-standards in lossless coding which cater for lossless compression applications. Most of these coders have a relatively simple structure in comparison to state-of-the-art lossy coders. Key features of image coding such as adaptivity are often neglected or not fully exploited.

Adaptive coders are capable of adjusting their characteristics to address the non-stationarities within images and hence improving the performance of the coder. The work reported in this paper is a part of our recent research conducted into adaptive lossless image coding and the possible improvements it can offer.

Figure 1 depicts a simple and popular paradigm for a lossless image coder. The coder consists of two major components, the predictor and the entropy coder, either or both of which can be made adaptive. In this paper, we will focus our attention on the predictor component of the coder.

\*The first author is supported in part by a CSIRO Division of Telecomm. and Industrial Physics postgraduate scholarship.

In many cases, the chosen predictor is a simple linear predictor such as those used in the JPEG standard [1]. It is also possible to calculate the optimum linear predictor (in a mean-squared-error sense) by solving the Wiener-Hopf equations which provides better performance compared to JPEG. This improved performance is obtained at the cost of a considerable amount of computation.

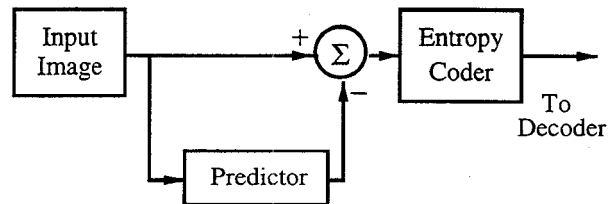


Figure 1. A DPCM-based lossless coder

Although the optimum linear predictor can be considered adaptive in the sense that it is defined for a particular image, it suffers from a number of drawbacks. The main drawbacks in using the optimum linear predictor are as follows:

- A significant amount of computation is required prior to coding.
- The predictor is still not truly adaptive. It does not adapt to the different characteristics of different areas within an image.
- The entire image must be known prior to designing the predictor (to calculate the autocorrelation coefficients).

There are two broad categories of adaptive prediction which can be applied to lossless image coding. Either, the coefficients of a single filter are constantly adapted to local characteristics within the image (eg. edges, textures, etc.), or a selection is made among a number of filters to choose the best predictor.

In other papers [4] [5], we have explored two scenarios where a finite number of predictors are selected one at a time for encoding particular regions within an image. However, in both papers, the proposed predictors require prior training and are computationally demanding. The computational expense is mainly due to the training and the decision of which predictor to use for a given pixel or group of pixels.

Alternatively, adaptive filtering techniques such as the LMS algorithm [2] may be used to make the linear predictor adaptive through adapting the filter coefficients. However, these adaptive predictors require additional calculations and if they are to adapt rapidly (as is often required in images), they are prone to becoming unstable.

In this paper, we propose a context-based adaptive predictor which overcomes some of the the abovementioned shortcomings. We propose a method of adapting a linear predictor's coefficients, which is both conceptually and computationally simple. By examining its immediate neighborhood within an image, the proposed predictor uses a "fuzzy" rule to adjust its coefficients. This allows the predictor to rapidly adapt to the local characteristics within an image, while remaining stable at all times.

## 2. THE CONTEXT-BASED ADAPTIVE PREDICTOR (CBAP)

The purpose of the predictor proposed in this paper is to adapt to the local characteristics of the areas within an image. Adaptivity is particularly important around edges and textures within an image where the largest prediction errors often occur.

In these areas a particular direction becomes dominant in the image (e.g. an edge), or we may also find that a particular pattern of pixels is repeated in a relatively large area (e.g. a texture). Linear predictors generally can not respond to these types of characteristics within an image.

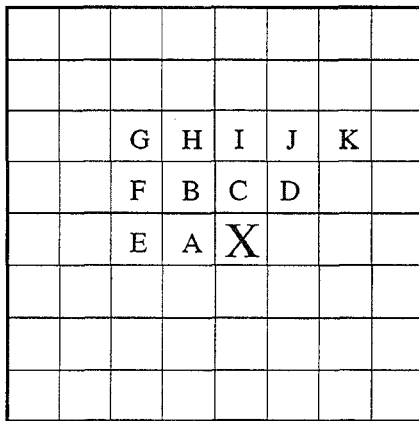


Figure 2. The causal context of pixel  $X$

The CBAP relies on a simple heuristic design which is capable of detecting repeated pixel patterns as well as edges within an image. In this predictor, predictions are made using a small neighborhood of previously transmitted pixel values. By keeping this neighborhood small, we force the predictor to rapidly adapt to the characteristics of that region.

Fig. 1 demonstrates the causal neighborhood of the pixel  $X$ . We use the notation  $I(X)$  to denote the intensity (pixel value) at pixel  $X$ . We intend to predict  $I(X)$  from the values of the 4 pixels  $A$ ,  $B$ ,  $C$  and  $D$ , which are available at both the encoder and the the decoder. We will refer to the pixels  $A$ ,  $B$ ,  $C$  and  $D$ , which we use to predict the pixel  $X$  as the context of pixel  $X$ . In this manner, we define

contexts of pixels  $X$ ,  $A, B, C$  and  $D$  as follows:

$$\begin{aligned} \text{Cxt}(X) &= \{A, B, C, D\} \\ \text{Cxt}(A) &= \{E, F, B, C\} \\ \text{Cxt}(B) &= \{F, G, H, I\} \\ \text{Cxt}(C) &= \{B, H, I, J\} \\ \text{Cxt}(D) &= \{C, I, J, K\} \end{aligned} \quad (1)$$

We hypothesize that if one of the pixels  $A, B, C$  or  $D$  has a context, which is similar to that of  $X$ , then that pixel and  $X$  are likely to be similar in value. For example, if pixel  $A$  has an identical context to that of pixel  $X$ , then it is likely that  $I(X)$  and  $I(A)$  are similar in value. Similarly, if there is less resemblance between the contexts of two pixels, then it is likely that the two pixels will be less similar in value.

In this paper, we use a fuzzy predictor to predict  $I(X)$  as a combination of  $I(A), I(B), I(C)$  and  $I(D)$ . For this, a membership function  $\mu(A, X)$  is defined such that it is large (approaches 1) when the contexts of  $A$  and  $X$  are similar, and small (approaches 0) when the contexts are less similar. We also define membership functions for other pixels in the context of  $X$  in a similar manner. With the aid of the membership functions, we define the fuzzy predictor for pixel  $X$  as follows:

$$\hat{I}(X) = \frac{\sum_{n \in \{A, B, C, D\}} \mu(n, X) I(n)}{\sum_{n \in \{A, B, C, D\}} \mu(n, X)} \quad (2)$$

where  $\hat{I}(X)$  is the predicted value of  $I(X)$ . Using the above predictor, the pixels whose context is more similar to that of pixel  $X$  will be allocated a larger value of  $\mu$  and hence will make a larger contribution to the predicted value of  $\hat{I}(X)$ . If the contexts are less similar, then their associated  $\mu$  will be smaller and hence their contribution to the total will also be smaller.

In order to define the fuzzy membership function, we first define the distances among the contexts as follows:

$$\begin{aligned} \text{Dist}(A, X) &= [I(E) - I(A)]^2 + [I(F) - I(B)]^2 + \\ &\quad [I(B) - I(C)]^2 + [I(C) - I(D)]^2 \\ \text{Dist}(B, X) &= [I(F) - I(A)]^2 + [I(G) - I(B)]^2 + \\ &\quad [I(H) - I(C)]^2 + [I(I) - I(D)]^2 \\ \text{Dist}(C, X) &= [I(B) - I(A)]^2 + [I(H) - I(B)]^2 + \\ &\quad [I(I) - I(C)]^2 + [I(J) - I(D)]^2 \\ \text{Dist}(D, X) &= [I(C) - I(A)]^2 + [I(I) - I(B)]^2 + \\ &\quad [I(J) - I(C)]^2 + [I(K) - I(D)]^2 \end{aligned} \quad (3)$$

The fuzzy membership function  $\mu(A, X)$  is defined as:

$$\mu(A, X) = \begin{cases} 1.0, & \text{if } \text{Dist}(A, X) < 1.0, \\ \frac{1}{\text{Dist}(A, X)}, & \text{Otherwise.} \end{cases} \quad (4)$$

Other membership functions are defined in an identical manner. The  $\mu$  values will range between 1 (strongest resemblance) and approximately  $2^{-18}$  (smallest resemblance).

Equations 2,3 and 4 completely define the predictor and are sufficient to make a predictor for use in an image coder. In a lossless DPCM coder, the predictions  $\hat{I}(X)$  are rounded off to the nearest integer before being subtracted from actual values.

This simple predictor is able to quickly detect a dominant direction in a texture or the direction of an edge. Although the fuzzy predictor is not optimal in any mathematical sense, we shall later demonstrate how it consistently outperforms the optimum linear predictor. This is solely due to the adaptivity of the fuzzy predictor.

### 3. RESULTS

The context-based adaptive predictor was implemented as a part of a lossless DPCM image coder (as in Figure 1) and tested on a number of monochrome images. The test images were 512x512 pixels in size and contained 256 grey levels.

For the purpose of comparison, the results from the CBAP are compared with those obtained from the optimum linear predictor (OLP) and the JPEG [1] image coding standard. We should note that the optimum linear predictor is only optimal in a Mean-Squared-Error (MSE) sense; while in this application we are interested in minimizing the entropy of the prediction residual. However, entropy-optimized predictors, require a very computationally intensive design and only result in modest gains over their MSE optimized counterparts. It is for this reason that MSE-optimized predictors were used for comparison.

There are seven predictors defined within the lossless JPEG standard, in these experiments predictor number 7 (refer to [1]) was used. This is a simple predictor with two coefficients of 0.5. Among the various JPEG predictors, this predictor often produces the best results.

An adaptive arithmetic coder [3] was used to encode the prediction residual. The results listed in Table 1 are all based on the file sizes produced by the coder.

From Table 1, it is evident that the coder using CBAP outperforms both the optimum linear predictor and the JPEG coder. This is particularly evident in images which contain areas which are highly directional such as the "barb" and "bank" images. In other images, the CBAP still outperforms the other coders by a significant margin.

Examining the average bit-rates, reveals that CBAP outperforms the JPEG standard and the OLP by 0.47 bpp and 0.3 bpp respectively. In particular, it is interesting to note that CBAP outperforms the optimum linear predictor by this margin without requiring any prior training as is the case for the OLP. The only two images from which the OLP outperformed CBAP were the "vegas" and "hat" images. This is possibly due to the absence of sharp edges and strong textures in these images. Those are the areas in which we expect CBAP to perform at its best.

The CBAP coder takes longer to encode an image than both the OLP and the JPEG coders. However, if we take into account the training time required to calculate the OLP coefficients, then the CBAP is by far the faster. As expected, the JPEG lossless coder runs in the shortest time. We should also note that the complexity of the CBAP is symmetric. That is, the CBAP coder and decoder run in

similar times.

### 4. CONCLUSIONS

The context-based adaptive predictor proposed in this paper provides a simple, yet effective method of incorporating adaptivity into a lossless image coder. The test results show that even though the CBAP requires no prior knowledge of the image it can perform as well as or often better than the optimum linear predictor. The advantages of this predictor are most evident when used for images which contain highly directional regions.

It should be noted that the CBAP is more computationally demanding than the fixed predictors used in the lossless JPEG standard. However in many cases, this drawback may be offset by the higher compression ratios offered by CBAP. We should also note that in situations where coding time is important, many of the CBAP operations can be parallelized to speed up the algorithm.

The encouraging performance of the CBAP algorithm on images which contain highly directional regions, also makes this algorithm an ideal candidate for compressing synthetic images (computer graphics, etc.) which contain many sharp edges and repeated pixel patterns.

Coded Image	Coding Method		
	CBAP	OLP	JPEG
bank	4.34	4.71	4.75
barb	5.06	5.27	5.52
couple	4.81	4.92	5.17
crowd	4.36	4.36	4.68
lax	5.82	5.89	5.93
lena	4.33	4.44	4.62
man	4.69	4.80	4.93
fruit	4.59	4.79	4.91
hat	4.27	4.15	4.46
vegas	4.65	4.60	4.81
woman1	4.80	4.96	5.04
woman2	3.51	3.53	3.68
boat	4.58	4.69	4.89
goldhill	4.90	4.91	5.05
zelda	4.03	4.11	4.29
Average	4.38	4.68	4.85

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

### REFERENCES

- [1] Gregory K. Wallace, "The JPEG still picture compression standard," *Comm. ACM*, vol. 34, pp. 30-44, April 1991.
- [2] B. Widrow, "Adaptive filters," *Aspects of network and system theory*, ed. R.E. Kalman and N. DeClaris, Holt, Rinehart and Winston, 1970, New York.
- [3] I.H. Witten, R.M. Neal, and G.J. Cleary, "Arithmetic coding for data compression," *Communications of the ACM*, vol. 30, pp. 520-540, June 1987.



Figure 3. Test images used as listed in Table 1 (Left to right, top to bottom).

- [4] F. Golchin and K.K. Paliwal, "Classified adaptive prediction and entropy coding for lossless coding of images," *Proceedings of the International Conference on Image Processing ICIP-97*, Santa Barbara, CA, October 1997.
- [5] F. Golchin and K.K. Paliwal, "Context classification and adaptive prediction for lossless image coding," *Proceedings of the IEEE Region 10 Annual Conference TENCON-97*, Brisbane, Australia, December 1997.