IMPROVING THE EFFICIENCY OF PREDICTIVE CODERS VIA ADAPTIVE MULTIPLE PREDICTOR COOPERATION

Huseyin Sar, Cihan Topal, Nuray At and Omer Nezih Gerek
Department of Electrical and Electronics Engineering
Anadolu University, Turkey
e-mail: {hsar, cihant, nat, ongerek}@anadolu.edu.tr

ABSTRACT

Due to the popularity of the prediction concept in time series analysis, predictive coding has been an attractive approach, particularly in lossless image compression. Utilization of prediction in time series not only makes use of residual encoding of the prediction error, but also describes and models the behavior of the underlying process. Unfortunately, this approach seems to have limited most of the scientists in the compression society to focus only to causal (or windowed) predictors, which are fine tuned to particular signal patterns. This work considers the fundamental formulation of finite extent data compression by making use of "adaptive multi-channel" prediction that is constructed by comparing prediction values of separate predictors (called, the multiple predictor cooperation). The deliberately generated channels are observed to have sharp error distributions with different bias centers. These biases are centered in a second pass, to produce plausible experimental predictive compression results.

Index Terms— Predictive coding, predictive error distribution, bias

1. INTRODUCTION

Predictive coding is one of the important tools in communication and data compression. Even though a deceleration has been observed in predictive image coding research since the adoption of the JPEG-LS standard, noteworthy developments have emerged recently.

The classical prediction approach is borrowed from time-series analysis. Predictive coders use neighborhood relations (correlation) between a sample of interest and its causal past, to remove the information redundancy. The prediction error, difference between the prediction and an actual sample value, is desired to be small. The "smallness" of the prediction error is typically measured in terms of the variance or the entropy of the error sequence.

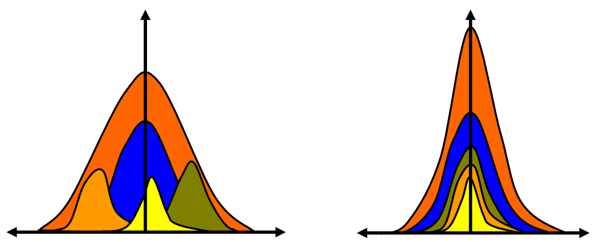
In [1], Motta et al proposed a single-pass Adaptive Linear Prediction and Classification algorithm, ALPC. In that algorithm, a pixel of interest is predicted by a weighted sum of its causal neighbors where the weights are chosen to minimize the energy of the prediction error inside a small prediction window. In [2], Li and Orchard suggested a least-square (LS)-based adaptive algorithm. Recognizing the edge-directed property (EDP) of the LS-based adaptation, they propose to perform LS optimization mainly around the edge areas (rather than performing for the whole image) thus reducing the computational complexity. In [3], Wu and Memon proposed a Context-based, Adaptive, Lossless Image Codec (CALIC) operating in two modes: binary and continuous. In continuous-tone mode, CALIC uses Gradient Adjusted Prediction (GAP) where local gradients of the intensity function at the current pixel are first estimated and a prediction is made according to these gradients. The local estimated gradients are used to determine edge structures as horizontal/vertical edge, sharp (or weak) horizontal/vertical edge, or smooth area. Another fundamental work on predictive image coding is Low Complexity Lossless Compression for Images (LOCO-I) [6] used in the JPEG-LS. LOCO-I uses Median Edge Detector (MED) as a predictor. In [5], a combination of GAP and MED is proposed and called Gradient Edge Detection (GED) predictor. We also note the recent study [4] by Wu et al, where the proposed minimum description length (MDL) –based adaptive predictor is empirically established to be the best among all when applied to lossless image coding.

Regardless of the numerical success of a particular prediction method mentioned above, it can be claimed that these methods utilize context optimizations or adaptations with respect to correlations or nonlinear correspondences. Therefore, the idea of prediction is still borrowed from time series analysis, where the predictor is sought as a model for the description of the signal.

Depending on the local context, each predictor (MED, GAP and GED) produces a prediction output. These predictors also flag the type of the context that they use. In the 2007 paper, the trivial case of adjusting the context depended prediction outputs according to the prediction error means of each context was considered. The motivation behind that means adjustment approach was the observation of different prediction error distribution structures, particularly their means for each context. For different predictors, the prediction output distributions have different context dependencies, depending on the predictor used. It is, therefore, a reasonable approach to pass the image through each of these predictors (MED, GAP and GED) and consider their context-dependent outputs for further post-processing. In this paper, the comparison of the context depended prediction outputs of each predictor is performed to split the output prediction error stream into, so-called, channels; hence the cooperation of the predictors. ıt is observed that the produced channels have good (sharp and narrow with small variance and entropy) distribution shapes that may enable good performance by the following entropy coders.

The idea presented in this work is inherited from the basic concepts of data compression, where causality, single context, data correspondence, etc. are not necessarily important. The ultimate aim is to have a list of prediction error samples, which form a sharp histogram/distribution (meaning low entropy or energy). For this purpose, we propose to form groups of prediction errors, coming from multiple prediction channels in an adaptive manner. MED [6] is a typical example of a multiple channel predictive coder, where the prediction output is conditioned into 3 cases.

In [7], Topal and Gerek proposed to handle such channels separately before collecting them into a single pool, which yielded an encouraging amount of histogram optimization, i.e., sharper histogram. Note that the channel histograms are individually sharp but since each has a different mean (bias) value, the combination of them results a wider distribution for the final prediction output. Consequently, when the channel means are compensated and hence aligned, the final prediction output would have a sharper histogram or smaller entropy as illustrated in figure 1.The idea is illustrated in Figure 1, where (a) shows histograms of generated channels with different mean values and of the total prediction output, and (b) shows mean-compensated individual channel histograms and the resulting narrower total prediction output histogram.



(a) (b)

Fig. 1. Histograms Of The Error Channels (a) Before And (b) After Bias Cancellation

The above observation is an example of how channels could be generated by comparing contexts. On the other hand, it is also possible to generate channels by comparing different predictors. In this work, we employ multiple predictors to obtain a multi-channel system. We compare outputs of the predictors and decide which predictors to cooperate for the purpose of splitting an error sequence into adaptive sub-channels, hence called "adaptive predictor cooperation". The generated channels usually exhibit different mean values with small variances, indicating the possibility for optimization, by aligning the histograms about a specific value. Clearly, the determination of the mean of each channel requires one complete pass of the prediction algorithm. However, once the means are determined, the mean compensation in cooperating channels can be done easily. During the first pass the whole prediction sequence is produced and saved so that on the second pass there is no need to reproduce the prediction errors, that just offsetting the saved prediction error sequence is the work of second pass. Therefore the second pass does not add an extra complexity. The overall complexity of the proposed method is not really more than the single pass of the corresponding prediction algorithm.

In this study, we improve the cooperative prediction idea introduced in [7, 8] in several ways: (1) the use of multiple predictors, (2) the use of more sophisticated cooperation strategy and (3) adaptive channel splitting stage. With the proposed approach, better entropy results are obtained in general compared to that of each employed prediction algorithms. For example, when the cooperated algorithm is MED, an entropy improvement of about 1.64% (on the average) is achieved via cooperation with other efficient predictors. Similar improvements are observed for other cases of *cooperations*.

2. PREDICTION ERROR OPTIMIZATION

Due to the unavoidable sub-optimality of any predictor, the error residue of the prediction still needs to be optimized. This becomes even more eminent at edge areas of an image, where mean-shifted error distributions are possible. In conditional (multi-channel) algorithms, error bias cancellation is an important problem – also due to the condition dependent (non)linear predictors. CALIC and LOCO-I encoders have predictors (GAP and MED, respectively) that suffer from the bias dissolve problem, too. In CALIC, this problem is remedied by estimating the error within each context and then adding these bias error values to the predicted value to get sharper distribution [3]. Similarly in LOCO-I, the bias parameter is estimated in the form of a normalization factor to cancel that parameter [6]. As mentioned before, these attempts consider the prediction phenomenon as a causal analysis tool, without even considering "calculation" of the bias instead of trying to "estimate" them. On the contrary, the proposed algorithm attempts the bias control problem by distributing the total error logically to bins (channels) each with sharp distributions and possibly different biases adaptively. The rest of the idea is to optimize total error by keeping and compensating-for the biases at each channel [8]. As a result, it is not necessary to *estimate* bias anymore.

4. MULTIPLE COOPERATED PREDICTORS

In this section, we review three predictors that we employ in our proposed architecture. Note that the proposed approach is actually predictor independent. Hence, the introduced predictors are used for illustration purposes.

4.1 MED

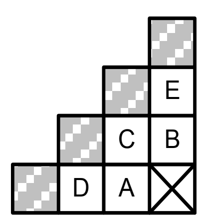
Median Edge Detector (MED) is one of the well-known prediction algorithms used in the JPEG-LS (LOCO-I) [6]. MED uses three causal neighboring pixels, performing a primitive test to determine the edge structure as vertical edge, horizontal edge, or smooth area, and the current pixel value is predicted accordingly.

4.2 GAP

Gradient Adjusted Prediction (GAP) is used in the commonly accepted benchmark coder CALIC [3]. GAP employs seven causal neighboring pixels to estimate gradients in the horizontal and vertical directions. The difference between gradients is then compared with empirically chosen thresholds to determine the edge structure as horizontal/vertical edge, sharp (or weak) horizontal/vertical edge, or smooth area. According to the edge structure, one of the predictors is assigned to make a prediction for the current pixel. For details of the algorithm, see [3].

4.3 GED

The Gradient Edge Detection (GED) predictor [5] can be seen as the combination of MED and GAP. Figure 2 shows the causal template and the GED algorithm.



Definition:

$$g_v = |C-A| + |E-B|$$

$$g_h = |D-A| + |C-B|$$

$$P = \begin{cases} A, & g_v - g_h > T \\ B, & g_v - g_h < -T \\ 3\dfrac{A+B}{8} + \dfrac{C+D+A}{12}, & else \end{cases}$$

Fig. 2. Causal Template And Ged Algorithm

Specifically, GED uses five causal neighboring pixels to estimate gradients in the horizontal and vertical directions. The difference between gradients is compared to a threshold value T to determine the edge structure as vertical edge, horizontal edge, or smooth area and the current pixel value P is predicted accordingly. In our experiments, we empirically choose the threshold value T as 22.

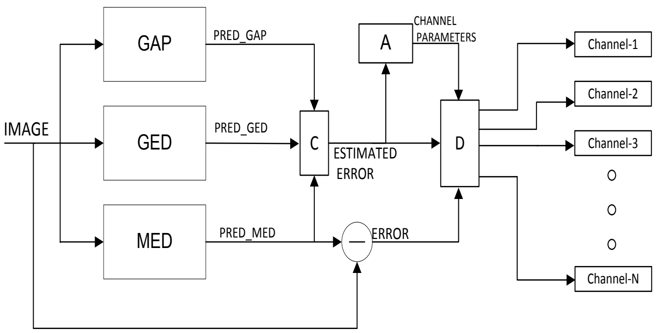


Fig. 3. Proposed Algorithm

6. PROPOSED METHOD

We propose an adaptive prediction error distribution sharpening strategy for predictive image coders as seen in Figure 3. In the first stage of the proposed algorithm, three predictors (GAP, GED, and MED) are run simultaneously and a predictor P_{COOP} that cooperates with MED is determined via The Multiple Cooperation Algorithm given in Figure 4 (C block in the Figure 3). For each pixel, difference between the predicted values produced by MED and P_{COOP} is calculated, called *estimated error*. Next, using the histogram of the estimation error, a number of channels and channel margins are determined adaptively in A block in Figure 3. In block A, by utilizing the estimated error histogram's mean; fair distribution of prediction error values into each channel is provided. The total number of prediction error values for each channel becomes almost fixed. Therefore, the effect of each error channel histogram to final distribution is compensated and sharper error channel histograms can be obtained. Note that the channel parameters (number of channels and channel margins) will be specific to a picture under study. Using the channel parameters and the error between MED and actual value, sub-channels are appointed (D block in Figure 3). In block D by comparing the difference between predictions of MED and P_{COOP} with channel margins, the current error value is being addressed to decided channel. Then, similar to the work in [8], biases in error sub-channels are determined and compensated.

In the cooperation idea [8], the cooperating predictors were chosen as MED and linear average predictor within the same structuring size. However, the linear average predictor is rather a primitive predictor choice for the cooperation. Moreover, utilization of multiple predictor choices was not implemented in that study. In this work, it was observed that the more successful the individual cooperating predictors are, the more satisfactory (sharp) error sub-channel distributions are obtained.

It must be noted that the proposed work is not a new prediction algorithm. It is an improvement tool to deliberately split the prediction error samples to adaptively constructed channels, which are expected to contain different biases, enabling the possibility of further performance improvement by compensating for the channel biases. It was experimentally observed that, by using successful predictors in the channel split selection phase (the cooperation), better overall optimizations could be achieved. Technically, the proposed improvement algorithm stays valid regardless of the prediction algorithm. So, even when a more successful predictor is encountered in the future, the proposed method can be incorporated for its further improvement. The entropy values of prediction residues with the above mentioned predictors are given in the experimental results section. According to the experimental results, the proposed approach is shown to outperform each individual predictor utilized herein.

On the prediction error stream rectification process, it must be rated that the context depended separation into different channels does not require additional tagging of each channel at the output bit stream. The decoder of the proposed prediction scheme is completely symmetric. The decoder uses past (previously decoded) image samples to give the same context and channel decision as the encoder and compensates according to the indicated mean shift. The only additional bits are, therefore, due to the "for once" representation of the shift amounts of each channel. For example, if there are ten channels, the total amount of overhead is 10 numbers (each being one byte) appended to the header. This case can be quantified with the case of "512 x 512 peppers" test image as follows. According to the entropy results in Sec. 8, this image is compressed to 147,850 bytes. With 10 more bytes, the output becomes 147,860 bytes. On the other hand, the closest compression is achieved by the GAP predictor, which compresses the image to 151,680 bytes. Clearly, the overhead is reasonable compared to the representation of the whole image and the algorithm does not suffer from context dilution.

8. EXPERIMENTAL RESULTS

Illustrations after the performance of the proposed adaptive multiple-cooperation method is demonstrated by error distributions and entropy results in comparison with the individual utilization of GAP, MED and GED. Some standard 512x512 test images are used. The entropy results are given in Table 1. As can be seen from the table, the proposed method almost always has lower entropy rates than GAP, MED, GED and Method in [8]. For instance, concerning the "peppers" test image, the entropy improvements with respect to GAP, MED, GED and Method in [8] are 2.53%, 6.47%, 3.14% and 2.32% respectively.

In Figure 5 the error images (Boats Test Image) of the GAP, GED, MED and our Proposed Method are given, where the smoothness of the proposed method can be seen.

Definition:

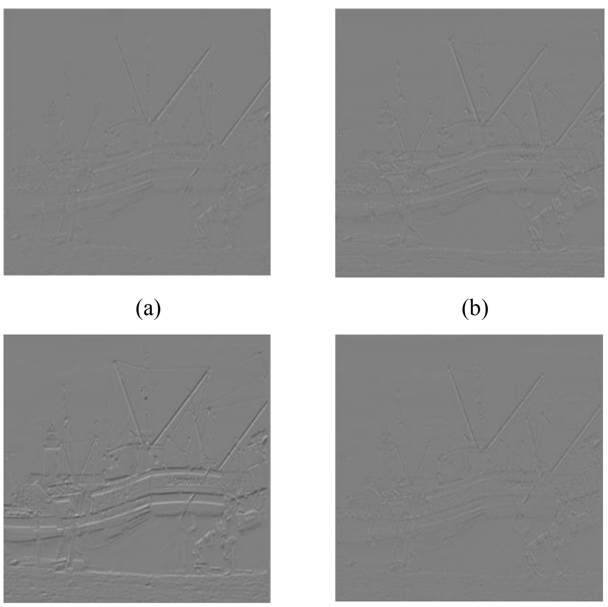
$$P_{COOP} = \begin{cases} \min(P_{GAP}, P_{GED}), & P_{MED} = \max(P_{GAP}, P_{GED}, P_{MED}) \\ \max(P_{GAP}, P_{GED}), & P_{MED} = \min(P_{GAP}, P_{GED}, P_{MED}) \\ P_{GED}, & P_{MED} = P_{GAP} \\ P_{GAP}, & P_{MED} = P_{GED} \\ \min(P_{GAP}, P_{GED}), & else \end{cases}$$

Fig. 4. The Multiple-Cooperation Algorithm

In Figure 6, error distributions of the separated 8 channels (prior to bias cancellation) are plotted. Note that number of channels "8" and the corresponding channel margins are chosen according to our adaptive channel parameter estimator. Separate biases (means of the distributions) are eminent. Combining these distributions in a central distribution (Figure 7) would clearly produce a sharper distribution than combining the channels without bias cancellation (mean centralization. The overall prediction error distributions of GAP, MED and GED are clearly not as sharp as the distribution of the proposed algorithmic improvement.

IMAGES	ENTROPY RESULTS OF PREDICTORS					
	Native Entropy	*GAP*	*MED*	*GED*	*Method in [8]*	*Multiple Predictor Cooperation*
Lena	7,44	4,40	4,54	4,68	4,53	4,40
Barbara	7,63	5,39	5,47	5,57	5,45	5,40
Sailboat	7,31	5,26	5,38	5,43	5,30	5,26
Peppers	7,59	4,74	4,94	4,77	4,73	4,62
Pentagon	6,65	5,22	5,32	5,45	5,27	5,22
Cameraman	6,85	3,73	3,55	4,33	3,57	3,55
Boats	7,03	4,37	4,40	4,73	4,35	4,30
Goldhill	7,47	4,84	4,87	5,23	4,85	4,82
Airplane	6,70	4,13	4,18	4,60	4,15	4,13
Couple	7,05	4,81	4,78	5,29	4,81	4,79
Baboon	7,35	6,20	6,24	6,45	6,23	6,21
Harbour	6,75	5,00	4,94	5,33	4,98	4,93
Average	**7,15**	**4,84**	**4,88**	**5,16**	**4,85**	**4,80**

Table 1. Self-Entropy Results (BITS PER PIXEL) Of Test Images (Size: 512x512)



(a) (b) (c) (d)

Fig. 5. Error Image of the "Boats". (a) MED, (b) GAP, (c) GED and (d) Proposed Method

9. CONCLUSION

An adaptive systematic error distribution sharpening strategy is proposed for predictive image coders. The proposed method compares the prediction error values coming from separate-but-simultaneously working predictors and provides this information as a decision criterion for splitting the error channel of one of these predictors. The achieved error sub-channels are observed to contain sharp-but-biased distributions. Consequently, by cancelling these biases and combining the error channels, the overall error distribution entropy is observed to improve.

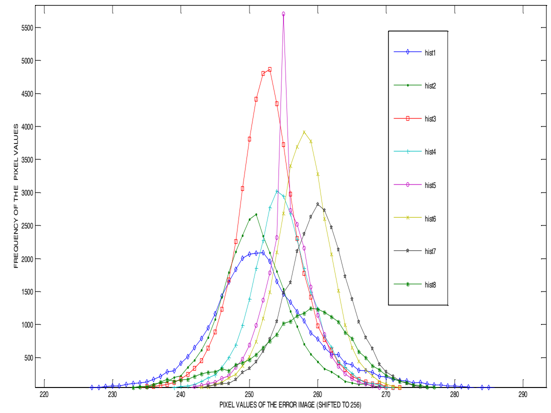


Fig. 6. Error Channels Distributions Before Bias Cancellation For The "Peppers" Test Image

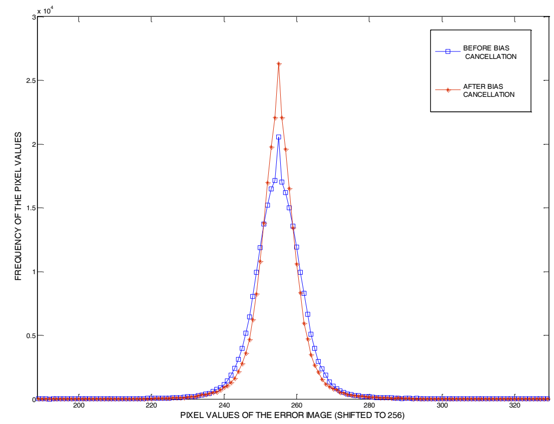


Fig. 7. Total Error Distributions Of The Proposed Method Before And After Bias Cancellation For The "Peppers" Test Image

12. REFERENCES

[1] G. Motta, J.A. Storer, B. Carpentieri, "Lossless image coding via adaptive linear prediction and classification," Proc. of the IEEE, vol.88, no.11, pp.1790-1796, Nov 2000.

[2] X. Li, M.T. Orchard, "Edge directed prediction for lossless compression of natural images," Int'l Conf. Image Processing (ICIP), vol.4 pp.58-62, 1999.

[3] X. Wu, N. Memon, "Context-based, adaptive, lossless image coding," IEEE Trans. Communications, 45(4), pp.437-444, Apr 1997.

[4] X. Wu, G. Zhai, X. Yang and W. Zhang, "Adaptive Sequential Prediction of Multidimensional Signals with Applications to Lossless Image Coding," IEEE Trans. Image Processing, vol.20, no.1, pp.36-42, Jan 2011.

[5] A. Avramovic; B. Reljin; "Gradient edge detection predictor for image lossless compression," Proceedings of ELMAR, pp.131-134, 2010.

[6] M.J. Weinberger; G. Seroussi; G. Sapiro; , "The LOCO-I lossless image compression algorithm: principles and tandardization into JPEG-LS,", IEEE Trans. Image Processing, vol.9, no.8, pp.1309-1324, 2000.

[7] C. Topal, Ö.N. Gerek, "Pdf sharpening for multichannel predictive coders," Proc. 14th European Signal Processing Conference (EUSIPCO), Italy, 2006.

[8] C. Topal and O. N. Gerek, "A Novel Lossless Image Compression Approach: Cooperative Prediction", Picture Coding Symposium (PCS 2007), Portugal, 7-9 Nov.2007