A NOVEL LOSSLESS IMAGE COMPRESSION APPROACH: "COOPERATIVE PREDICTION"

Cihan TOPAL and Ö. Nezih GEREK

Anadolu University, Department of Electrical and Electronics Engineering,
Eskisehir, TURKEY.
e-mail: {cihant,ongerek}@anadolu.edu.tr

ABSTRACT

Conditional predictive image coders (such as LOCO, CALIC, etc.) split the prediction rule into logical cases (channels) and produce prediction residuals for each case. It is a known fact that the distributions of these separate channels usually exhibit sharp, but mean-shifted shapes. If the mean-shift amount for each channel is determined and compensated for, the overall prediction error provides smaller entropy with a sharper distribution. In this work, several prediction rules are tested for obtaining sharp and possibly mean-shifted or skewed individual prediction channel outputs. The overall prediction output was not considered as the optimization criteria. By compensating for the shifts of each channel mean, very sharp and symmetric distributions are sought at each channel, so that the combination of these channels provides an overall sharp prediction error distribution. It is shown that the proposed method provides better compression results than the celebrated LOCO which is a well-known and efficient lossless compression algorithm.

Index Terms— Distribution enhancement, Lossless predictive image coding, prediction error classification.

1. INTRODUCTION

In predictive image coders that utilize time varying and image dependent predictors, the number of prediction error channels can be considered as the number of cases in the prediction stage. LOCO is a good example for such coders[1][2][3] where the input is conditioned to have three structural possibilities, producing three separate prediction error channels [2]. It has been observed that, if each of these prediction channels are analyzed in detail, a pronounced amount of mean-shift may be observed in their individual distributions [1]. These shift amounts usually cancel and add up to a symmetric distribution, once the prediction error samples are considered all together. In [1], it was shown that compensating these shifts within the prediction channel greatly improves the entropy of the overall prediction error. That was a simple addition to the LOCO encoder with the same number of prediction channels. However, construction of other multi-channel predictive coders is possible with the same idea. Starting from this idea, several prediction condition pairs are constructed and tested. These conditions are desired to produce several channels, all with a possible amount of mean-shift, but with as small variance as possible. If each channel output exhibits a sharp distribution, their summation after individually compensating for the shift would produce a total prediction error with smaller entropy and variance. Obviously, the shift compensation amount must be made available to the decoder using extra bits. On the other hand, since the number of channels are relatively few, the extra bits used do not add up to a disturbing amount in the overall coded bit stream.

2. PREDICTION ERROR SEPARATION

The critical step for the success of mean-shift compensation is the logical separation of prediction error channels such that each channel distribution is sharp. Usually, if the whole prediction error signal is considered, its distribution is symmetric as depicted in Figure 1. The distribution may look plausible with a good symmetry and narrow variance. However, such a distribution normally consists of a combination of several signal distributions coming from different prediction channels.

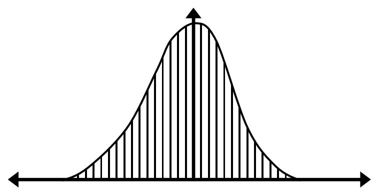
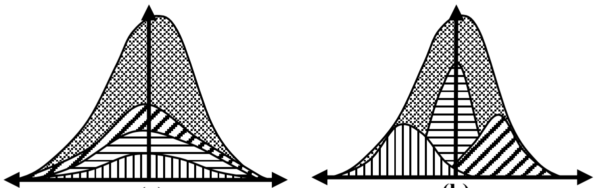


Figure 1. Cumulative prediction error distribution

The sub-distribution combinations may be exemplified in Figure 2(a) and 2(b). In these figures, the resulting error distribution is composed of error distributions coming from three different channels. Clearly, if the channel error distributions are either as in Figure 2(a) or as in 2(b), the resulting distribution will be the same. On the other hand, the proposed improvement is possible only if the resulting prediction error is generated from channels resembling a situation as in Figure 2(b). In such a situation, the individual prediction error channels have smaller variance, however they exhibit some amount of shift in the mean. In this work, the desired cooperative coder should have channel distributions in the form of Figure 2(b).

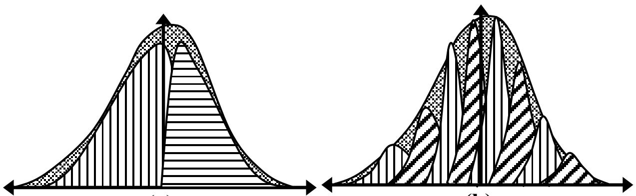
Another issue of the multichannel predictive coding is the number of channels for prediction. The optimum situation corresponds to many prediction channels with as



(a) **(b)**

Figure 2. (a) Inefficient**, (b)** efficient error separation.

sharp error distributions, as possible. An improvement situation is illustrated in Figures 3(a) and 3(b).



(a) **(b)**

Figure 3. (a) 2-channel efficient separation,
(b) 8-channel efficient separation**.**

In the 2-channel case, the best situation corresponds to separate channels with variances equal to half of the cumulative distribution (as in Figure 3(a)). The shift alignment of these two distributions certainly provide some improvement, however the gain would be marginal as compared to the 8-channel case, where the variance may improve up to 1/8 of the original. As a result, using many prediction channels is a desirable property. The critical issue is: how do we increase the number of prediction channels while keeping their output distributions as sharp as possible?

3. MULTIPLE CHANNEL CONSTRUCTION

Several logical cases may be built for decomposing the total prediction error distribution. In [1], the prediction channels were taken as the celebrated LOCO [2] method. LOCO is a simple but reasonable space varying prediction method that splits the image pixel prediction rule into 3 logical cases corresponding to the diagonal variation of the image. Considering the local template of an image as in Figure 4, LOCO predicts X as

1) **Max(B,C) if both B and C are greater than A (corresponding to a -45° gradient),**
2) **Min(B,C) if both B and C are greater than A (corresponding to a 135° gradient), and**
3) **B + C - A if no gradient is detected.**

A	B
C	*X*

Figure 4. LOCO prediction template.

In above prediction channels, LOCO tries to obtain the "best" prediction under +/- 45^0 gradient transitions.

It has been observed that the overall optimization of the prediction channels is not particularly important if the means and skews of the prediction channels could be measured separately. In [1], for instance, it was shown that the three LOCO prediction channels (specifically channel 1 and 2) actually provide better variance figures if observed separately. By simply compensating for the shifts in the channel distributions and combining the compensated prediction error, the resultant prediction error was observed to have a smaller variance and smaller entropy. The problem about working on LOCO was its relatively marginal improvement. This was mainly because of the "already" optimized structure of LOCO which was not deliberately designed for sharp error distributions in "each" channel.

In this work, the idea of compensating for the shifted means at each prediction channels is extended beyond the available and popular predictive coders. As an immediate example, the case (hence channel) numbers of LOCO can be increased to 9. The original LOCO does not consider the ranking of B and C once they are either smaller or larger than A. Therefore, a further ranking check can be incorporated, and the number of channels can be increased to 9. The original LOCO algorithm and the extended (9-channel) algorithm are described as pseudo-codes in Figures 5(a) and 5(b), respectively. It must be noted that the extended algorithm already includes the cases of the original LOCO, therefore its performance would certainly be above the original.

The compression test results for these two cases are presented in Table 1. The percentages are standing for the entropy improvements over the original LOCO algorithm.

```
if (A is Max)
   prediction_1;
   err_channel_1;


else if (A is Min)
   prediction_2;
   err_channel_2;


else
   prediction_3;
   err_channel_3;
```

(a)

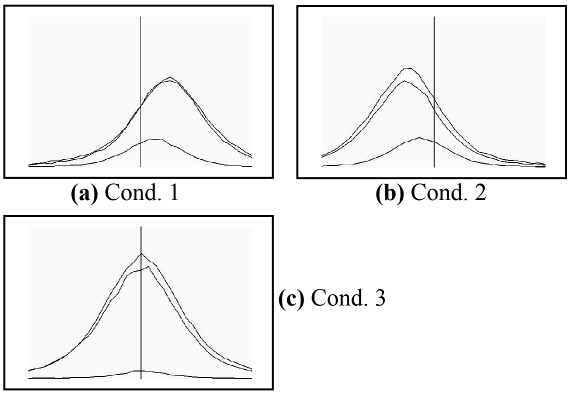
```
if (A is Max)
   prediction_1;
   if (B>C)
      err_channel_1;
   else if (B<C)
      err_channel_2;
   else  'means (B=C)
      err_channel_3;

else if (A is Min)
   prediction_2;
   if (B>C)
      err_channel_4;
   else if (B<C)
      err_channel_5;
   else  'means (B=C)
      err_channel_6;

else
   prediction_3;
   if (B>C)
      err_channel_7;
   else if (B<C)
      err_channel_8;
   else  'means (B=C)
```

(b)

Figure 5. (a) 3-channel LOCO algorithm,
(b) its 9-channel extension.

Although the 9-channel extension of LOCO improves the overall compression, the gain is marginal. This is mostly due to the fact that the mean shifting improvement is possible only in channels 1 and 2, and channel 3 is almost always symmetric (see Figure 6). The gradient logics of channels 1 and 2 consistently underestimate or overestimate the true value, yielding a bias. In channel 3, such bias does not exist. The unbiased characteristic is preserved even when the 3rd condition is split further.

This observation indicates that, if bias cancellation is the main issue for distribution sharpening, consistently biased prediction conditions may be deliberately used. The biased conditions must, however, provide a sharp error distribution that may be optimized by measuring the bias, and compensating for the bias before combining the total prediction error. It must be noted that, the bias compensation is a reversible operation. In lossless compression, if the pixels are reconstructed in a raster scan order up to a particular pixel of interest, the prediction rules work symmetrically for the decoder, and the decoder selects exactly the same rule as the encoder. If the compensation amounts are transmitted to the decoder side for each channel, the decoder can add the same compensation amount to the decided channel. This is, obviously, not possible in lossy compression where the prediction errors (hence the reconstructed pixels) vary, and that variation may cause the decoder to select an erroneous prediction channel which is not necessarily the same as the encoder side. To avoid this case, only lossless compression is assumed.



(a) Cond. 1 **(b)** Cond. 2 **(c)** Cond. 3

Figure 6. Sub-distributions of 9 channels of Peppers test image.

4. COOPERATIVE PREDICTION

The best (sharpest) array of prediction channels can be constructed by using the actual values of the pixels to be predicted. In that way, the channel number would be, say, 256, and each channel would produce a prediction error with zero variance. This is obviously pathological, since the value of the pixel is unknown to the decoder. At this point, the idea of prediction cooperation comes to ones mind: "If the true value of the pixel is not available, then one should try to estimate the value using another predictor, and make the prediction error variance smaller." The proposed cooperative prediction works as follows. The prediction template is used in two different (linear/nonlinear) predictors. These two predictors provide different prediction values. The difference between these two prediction values can be considered as an "indicator" about which channel should be selected for this particular case. In a sense, this difference is a prediction of prediction error.

The separation of the channels is selected according to the quantized values of the difference between two separate predictions. Since the two predictors cooperate in the determination of the channel splitting, the name for the algorithm was selected as cooperative prediction.

The cooperating predictors should be statistically different in their algebraic expressions. In this way, the difference between the two predictions contains the information regarding the bias of the prediction of one of the predictors. Experimentally, we have observed that if a nonlinear predictor (such as ones in LOCO) is accompanied by a static linear prediction filter, their output difference produces a useful value that can be used in splitting to distinct prediction channels.

As a simple example for a cooperative prediction algorithm, we can describe the cooperation of the channel-3 predictor of LOCO with an averaging predictor. The channel-3 predictor predicts the value as P1 = B + C - A for a template given in Figure 4. The averaging predictor predicts as P2 = (A + B + C) / 3. The difference between the two predictions (P1 and P2) provides the channel splitting rule as listed in Figure 7.

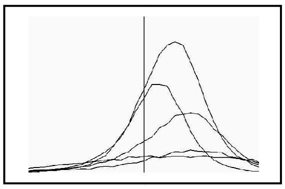
```
P1 = B + C - A;  'real prediction
P2 = Int((A+B+C)/3);'co-prediction
App_Err = P1 - P2;
if (App_Err<-6)
        err_channel_1;
else if (App_Err<-3)
        err_channel_2;
else if (App_Err<0)
        err_channel_3;
else if (App_Err<3)
        err_channel_4;
else
        err_channel_5;
```

Figure 7. A pseudo representation of a cooperative prediction algorithm.

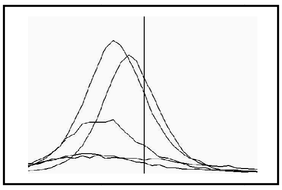
In this figure, 5 channels are obtained. In each channel, P1 may be accumulated and the bias for P1 within each channel may be calculated and compensated separately.

The idea of this predictive coder can be explained as follows: If the difference between the two predictions is small, that corresponds to a homogeneous region and either of the predictors accurately estimates the true value. Conversely, if the difference is high, that means an outlier value is encountered. By quantizing the difference, the localization of the outlier value can be separated. The negative and positive differences are observed to correspond to mean biases with different signs. Since the algorithm is lossless, the calculated prediction bias for, say, P1 can be obtained and added to the encoded bit stream. The decoder can use the same algorithm and previously determined and recorded bias amounts to perfectly reconstruct the image.

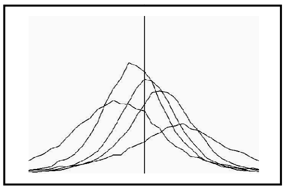
Nesting of the cooperations is also possible. In the following example, the three LOCO conditional predictors are cooperated with the linear averaging predictor. In each case, the prediction splits the channel into 5 more channels. As a result, a total of 15 channels are obtained. In Figure 8, these channels are presented in groups of 3 corresponding to the original LOCO conditions.



(a) Cond. 1



(b) Cond. 2



(c) Cond. 3

Figure 8. 15 channel prediction error distributions for image "Peppers"

5. EXPERIMENTAL RESULTS

The significant difference between Figure 6 and Figure 8 is the decomposition of the 3rd channel in Figure 8. Here, the symmetric distributions of the 3rd channel are successfully decomposed into biased (but sharp) distributions, each of which can be separately compensated to form an overall efficiency.

The above apparent efficiency is tested with several test images. The comparisons are made between the original LOCO and the bias compensated versions of original LOCO (3 channels), the logical splitting of LOCO (9 channels as in Figure 5), and the cooperative algorithm proposed here. The entropy reductions are given as percentages in Table I.

Table 1. Entropy improvements over the LOCO

Image	3-Channel	9-Channel Logical	15-Channel Cooperative
Peppers	% 2,71	% 2,73	% 4,13
F-16	% 0,00	%-0,03	% 1,05
Lena	% 0,30	% 0,37	% 0,33
Soundbarrier	% 0,00	% 0,16	% 1,20
Boats	% 0,60	% 0,44	% 1,23

The table indicates several observations. The first observation is; compensating for the mean shift biases in separate channels almost always improves the compression performance. Second, as the number of channels increase, the bias compensation efficiency becomes more pronounced. And third, the cooperation of two different predictors (deliberately selected as linear and nonlinear) provides the best improvements.

To test the efficiency of the proposed algorithm, a set of virtual laboratory environment comprising of three executable programs was developed in [5]. In these programs, results are compared to the LOCO algorithm in terms of entropy. The programs also provide statistical parameters of the conditional prediction outputs using a graphical user interface. Obviously, practical compression of the prediction residuals require optimization and very efficient programming of entropy coders, therefore it is left to be beyond the scope of this work which just illustrates the improvement in terms of entropy.

5. CONCLUSIONS

In this work, a lossless predictive coding approach based on splitting the prediction rule cases into multiple channels is presented. Unlike the classical multiple channel predictive coders that construct their predictors with logical contexts, the proposed technique does not try to obtain an overall prediction error sequence with low entropy. In this aspect, the paper argues that the concept of "contexts" in conditional predictive coders does not necessarily contribute well to the sharpness of the individual channel outputs. Conversely, contexts ensure a symmetric prediction error for the combined output. Without this symmetry constraint, it was observed that rules that do not depend on contexts occasionally provide sharper – but biased – conditional outputs. The prediction rules, shift determination, and compensation of them using extra bits in the compressed bit stream are addressed in this paper.

By the way, the channels are constructed so that each may have a very sharp prediction error distribution with a possible amount of bias.

By measuring the bias for the image to be compressed at each channel, the bias amounts can be compensated for. Obviously, the bias amounts are specific to the image, and they must be incorporated into the coded bit stream for perfect reconstruction. However, since the number of channels is relatively few, this overhead is not noticeable in the result, i.e. just one byte for each channel. Experimental results indicate that two statistically different predictors and their prediction errors can be compared to each other to provide a measure for channel splitting in an efficient manner. This indicates that the improvement is not due to the structure or the context of the predictors, but rather due to separating the prediction space to as many sub-channels with sharp residuals, as possible. By using such cooperating predictors in several sub-channels, the entropy of the encoded bit stream was observed to be improved up to 4% as compared to the successful and celebrated lossless predictive coders such as LOCO. Exhaustive tests for several types of predictors for cooperation and development of a practical executable coder remains as promising future works.

10. REFERENCES

[1] C.Topal and Ö.N.Gerek "PDF Sharpening for Multichannel Predictive Image Coders" EUSIPCO'06 Florence, Italya, 2006.

[2] M.J. Weinberger, G. Seroussi and G. Sapiro, "The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS", IEEE Trans. Image Proc., Vol.9, No.8, pp.1309-1324, August 2000.

[3] X. Wu, N. D. Memon, and K. Sayood, "A Context Based Adaptive Lossless/Nearly Lossless Coding Scheme for Tore Images". ISO Working Document. ISO/IEC SC29/WG1/N256, 1995.

[4] X. Wu and N. D. Memon. "CALIC – A Context Based Adaptive Lossless Image Coding Scheme". IEEE Transactions on Communications, 45:437 – 444, May 1996.

[5] Test applications are available at http://www.mm.anadolu.edu.tr/~cihant/coopre.