ADAPTIVE COMBINATION OF LINEAR PREDICTORS FOR LOSSLESS IMAGE COMPRESSION

GUANG DENG,* HUA YE AND LAURENCE W. CAHILL

Department of Electronic Engineering, La Trobe University

Bundoora,Victoria 3083, Australia

*Corresponding author. Email: d.deng@ee.latrobe.edu.au Phone: +61 3 9479 2036 Fax:+61 3 9471 0524. This research was supported by an Australian Research Council small grant.

Abstract

Lossless image coding is an essential requirement for medical imaging applications. Lossless image compression techniques usually have two major components: adaptive prediction and adaptive entropy coding. This paper is concerned with adaptive prediction. Recently, several researchers have studied prediction schemes in which the final prediction is formed by a combination of a group of sub-predictors. In this paper, we present an overview of this new type of prediction technique. We show that the basic principle of adaptive predictor combination has been extensively studied and applied to many science and engineering problems. We then describe our combination scheme which is based on the estimation of the local prediction error variance. Experimental results show that the compression performance of the algorithms that employ this new type of predictor is consistently better than that of state-of-the-art algorithms.

1 INTRODUCTION

In many medical imaging applications, it is essential that no information is discarded. Hence there is considerable interest in the development of efficient lossless image coding schemes. Lossless image compression techniques usually have two major components: adaptive prediction and adaptive entropy coding. This paper is focused on adaptive prediction. For more general discussions of lossless image compression techniques the reader is referred to papers by Memon *et al*[17] [19].

The state-of-the-art techniques first predict the grey level of the pixel using an adaptive predictor, then encode the prediction error by a context-based entropy coder. LOCO [32] has achieved the best compromise between compression performance and computational complexity. CALIC [33] has been the benchmark technique in terms of compression ratio for the past few years, and its computational complexity is also relatively low. The adaptive predictors used in LOCO and CALIC are simple. Their performance is thus limited.

Current lossless image coding research can be divided into two categories. The aim of the first category is to achieve better compression performance than CALIC and LOCO while keeping computational complexity reasonably low. The aim of the second is to determine the ultimate compression that can be achieved regardless the computational complexity.

Our recent research belongs to the first category. We proposed a predictor [6] which adaptively combines a group of fixed sub-predictors. Each sub-predictor uses some or all of four causal neighbouring pixels. The setting of sub-predictors is based on the correlation between the current pixel and its neighbouring pixels. The combination coefficients for the sub-predictors are determined by robust estimates of their mean square prediction errors. We have shown that when coupled with a context-based adaptive arithmetic coder, the performance of the proposed algorithm is better than CALIC and LOCO. In the meantime, other researchers [25] [13] have also reported prediction schemes that use combinations of sub-predictors. Their results are also similar to or better than CALIC.

In this paper, we present an overview of these adaptive predictor combination (APC) techniques. From a signal processing point of view, APC has been successfully used in image processing and adaptive estimation. However, its application in image coding has not been extensively investigated. Variants of APC have also found wide applications in a variety of other fields including social and economic forecasting. From a modeling point of view, this approach builds up an adaptive model by combining simple fixed models. It also overcomes the model uncertainty problem associated with model selection approaches based on certain criteria. A generalization of adaptive predictor combination, the Bayesian Model Averaging, which has been extensively studied in statistics, is yet to be applied to signal processing problems.

This paper shows that the APC techniques offer considerable performance advantages over other popular prediction techniques for lossless image coding. Our experimental results have also confirmed the effectiveness of APC.

Section 2 presents a brief review of adaptive predictor combination research in a multi-discipline perspective including: signal processing, adaptive control, statistical science, and economic forecasting. Section 3 presents our proposed algorithm formulated in a least mean square error framework. In Section 4 a comparison of a number of state-of-the-art lossless image coding algorithms is presented. Conclusions from this work are summarized in Section 5.

2 ADAPTIVE PREDICTOR COMBINATION

In this section, we briefly describe recent research in adaptive predictor combination for lossless image coding and its relationship with image filtering. This is followed by a brief introduction to the research and applications of prediction combination in other disciplines.

2.1 Adaptive predictors

Since images are normally non-stationary, image predictors must be adaptive so that they can change according to the local characteristics. Adaptation prediction can be performed in a number of ways. One way is to use a fixed predictor structure such as a linear combination of the causal neighboring pixel values and to change the coefficient for each pixel adaptively. The coefficients can be calculated based on the local gradients [24] [19]. They can also be calculated by using the least-mean-square (LMS) algorithm [4].

Another way is to have a set of simple and fixed sub-predictors and adaptively select one sub-predictor as the final predictor. The median edge detection (MED) predictor in LOCO and the gradient adjusted predictor (GAP) in CALIC are two examples of this category. In a comparative study [18], the MED predictor was reported to have the best performance among all submitted JPEG-LS proposals.

Still another way is to adaptively combine a subset or the whole set of sub-predictors to form the final predictor [25] [26]. This is an extension to the first category. The sub-predictors are usually simple fixed. They can be linear or nonlinear. We refer to this type of prediction scheme as the adaptive predictor combination (APC) technique. We will concentrate on this technique in the rest of this paper.

2.2 Predictor combination in image coding

When we combine a group of predictors, an intuitive way to determine the coefficients is to "penalize" predictors which result in large prediction errors for causal neighboring pixels. In [25], Seemann and Tischer proposed an algorithm as follows.

Refer to Figure 1, let $x(n)$ be the pixel to be predicted and $x(n-k)$ ($k = 1, 2, \ldots, N$) its causal neighboring pixels. Further let $p_j(n-k)$ ($j = 1, 2, \ldots, M$) be the jth sub-predictor for pixel $x(n-k)$. The penalty term for the jth sub-predictor is calculated as:

$$G_j = \sum_{k=1}^{N} \left| x(n-k) - p_j(n-k) \right| . \tag{1}$$

x (n-3) *NW*	*x (n-2)* *N*	*x (n-4)* *NE*
x (n-1) *W*	*x (n)*	

Figure 1: An illustration of the causal neighbouring pixels used in the combination of sub-predictors. $x(n)$ is the current pixel to be predicted.

The prediction for the current pixel is then given by:

$$P(n) = \frac{1}{D} \sum_{j=1}^{M} \frac{p_j(n)}{G_j}, \tag{2}$$

where $D = \sum_{j=1}^{M} \frac{1}{G_j}$ is the normalization factor. In equation (1) the sum of the absolute prediction errors for each sub-predictor is calculated over a small neighborhood. This quantity is used as a localized measure of the relative performance of the sub-predictor. An extensive test of the above algorithm [25] shows that it outperforms GAP, MED and two other fixed predictors in terms of the entropy of the prediction errors. When it is coupled with context-based error feedback and entropy coding, its compression performance is also better than that of LOCO and almost the same as that of CALIC. This predictor has also been applied to color images and audio signals. This work has been extended to an algorithm called history-based blending (HBB) [26] of sub-predictors. HBB uses an adaptive predictor which is a linear combination of a set of fixed predictors. The combination coefficients are determined by the Least Squares approach. The compression performance of HBB is better than that of CALIC for some images [26].

Lee [13] recently proposed an adaptive predictor in which the causal neighbouring pixels of the current pixel are combined with an LMS-based linear predictor. The combination coefficients are determined by using a Bayesian approach. More specifically, each neighbouring pixel and a linear predictor are treated as sub-predictors. The prediction error of a sub-predictor is modelled by a Laplacian distribution. The posteriori probability is used as the combination coefficient. Lee has also demonstrated that the prediction performance of combining a group of sub-predictors is better than that of selecting one of the sub-predictors.

Predictor combination can be regarded as building up an accurate model from a set of simple models. This has recently been addressed by some researchers. Popat and Picard [23] proposed a method that combines a number of probability mass functions (PMF) conditioned on a respective small neighborhood to form an estimate of the PMF conditioned on the union of all small neighborhoods. Zhang [37] proposed a number of algorithms that build a composite model from a few given ones. The TMW algorithm [20][21]calculates the probability distribution of the current pixel (CP) by combining the probability distributions of the CP conditioned on a set of predictors.

Forming a new predictor by combining several predictors is closely related to image filtering where outputs from different filters are combined. For example, the inverse gradient filter [31] is similar to the predictor used by Roos *et al* [24]. The fast image enhancement algorithm [3] is similar to the predictor of Seemann and Tisher [25]. The counterpart of our proposed predictor using a group of simple fixed sub-predictors (presented in section 3) is the optimal edge-preserving hybrid filter [30].

2.3 Predictor combination

Generally, if there are N possible predictions of an event, then one can choose to combine them or to select the "best" of them as the ultimate prediction. In 1969, Bates and Granger [1] pioneered the development of combining forecasts. Five different combination methods were described in [8]. Since then many papers have been published on this subject [5][9]. Recently, Palm and Zellner have studied issues related to the relative merit of combination and selection [22].

The combination of predictions based on the Bayesian framework has been actively studied in recent years by statistics researchers [10] [27]. In their research, each predictor is regarded as a statistical model. A model is usually based on certain assumptions. If these assumptions are correct, then the model will produce an accurate prediction. Otherwise, the model may produce an inaccurate prediction. Since whether or not the assumptions are correct (or to what extent they are correct) is usually not known for non-stationary signals such as images, therefore there is uncertainty about the model. Using a model from a group of available models is a risky process, because the model uncertainty is ignored.

The Bayesian Model Averaging (BMA) provides a coherent mechanism to account for this problem. According to BMA, given the data D and a group of models M_k, the Bayesian estimate θ_B of a parameter θ is determined by:

$$\theta_B = E[\theta|D] = \sum_{k=1}^{K} \theta_k Pr(M_k|D) \tag{3}$$

where $\theta_k = E[\theta|M_k]$. It has been demonstrated that BMA provides better predictive ability than any single model [10].

Predictor combination has also been used in other science and engineering disciplines. For example, it has been used in supervised learning [11], source coding [28], predictive VQ for image coding [29], and Kalman filtering [2] [15][16].

2.4 Summary

In summary, we have shown that APC is an emerging prediction technique for lossless image coding. APC is also closely related to prediction combination theories and algorithms which have been well established in forecasting and statistics research. Therefore, it is expected that the application of these theories and algorithms should provide new and potentially powerful tools for image compression.

3 THE ADAPTIVE PREDICTOR COMBINATION BASED ON VARIANCE ESTIMATION

3.1 Problem formulation

The proposed prediction of a pixel $x(n)$ is a linear combination of a set of sub-predictors:

$$P(n) = \sum_{j=1}^{M} \alpha_j(n) p_j(n) \tag{4}$$

where $\alpha_j(n)$ are the coefficients and

$$\sum_{j=1}^{M} \alpha_j(n) = 1. \tag{5}$$

The prediction error is given by

$$
\begin{aligned}
e(n) &= x(n) - \sum_{j=1}^{M} \alpha_j(n) p_j(n) \\
&= \sum_{j=1}^{M} \alpha_j(n) \left(x(n) - p_j(n) \right) \\
&= \sum_{j=1}^{M} \alpha_j(n) e_j(n) \qquad (6)
\end{aligned}
$$

where $e_j(n) = x(n) - p_j(n)$ is the prediction error for the jth sub-predictor. Therefore, the combination of sub-predictors is equivalent to the combination of their respective prediction errors. This justifies the intuitive approach used by Seemann [25] to assign a smaller weight to a worse sub-predictor.

The problem can be formulated as: given a group of models (sub-predictors) and the previously processed pixels, determine the prediction for the current pixel. Using a linear combination of the sub-predictors, the problem becomes how to find a set of coefficients based on certain criteria such as minimizing the mean square error. There are at least two ways to tackle the problem:

1. One way is to determine a set of coefficients for a block of an image such that the mean square prediction error of the block is minimized. The same set of coefficients are then used for the prediction of each pixel in the block. The coefficients are block dependent and must be sent to the decoder as side information. The size of block should be properly chosen such that the stationarity assumption holds.

2. Another way is to determine a set of coefficients for each pixel based on the minimization of the mean square prediction errors over a block of its causal neighbouring pixels. This block will be referred to as the training block shown in Figure 2 in following discussions. This method does not assume any optimum property. It is based on a simple assumption that if a predictor performs well for the training block, then it should also perform reasonably well for the current pixel. Using this method, the coefficients are truly pixel adaptive, and it is not necessary to send them to the decoder. Similar methods have been used in the Least Squares based-linear predictor design and have been applied to lossless image coding despite its computational complexity [12] [14] [34][35] .

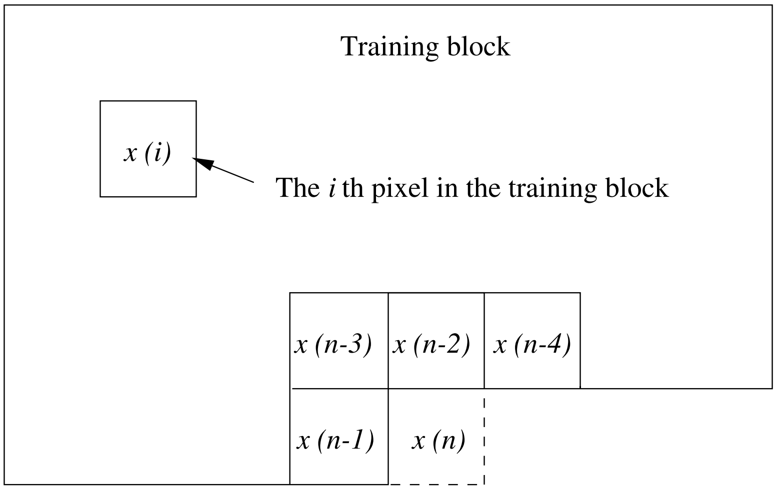

Figure 2: The training block is a block of causal neighbouring pixels of the current pixel.

In the following, we present the development of an algorithm to determine the predictor coefficients for the current pixel based on the above method. We use the Lagrange multiplier technique:

$$F = \sum_i e^2(i) + \beta \left(1 - \sum_{j=1}^{M} \alpha_j(n) \right) \tag{7}$$

to minimise the mean square error of the predictor (4) over the training block with condition (5). In equation (7), $e(i) = x(i) - P(i)$ is prediction error for the ith pixel in the training block and β is a constant to be determined. The coefficients can be found by solving the following set of equations:

$$\frac{\partial F}{\partial \alpha_j(n)} = 0,\ j = 1, 2, \ldots, M, \tag{8}$$

$$\frac{\partial F}{\partial \beta} = 0. \tag{9}$$

Assume that the prediction error of every sub-predictor has a zero mean, i.e., $\sum e_j(i) = 0$ ($e_j(i) = x(i) - p_j(i)$). In the following, the location index i is omitted to simplify the presentation. Let C be an $M \times M$ matrix whose elements are $\sigma_{jk} = \sum\sum e_j e_k$. Each diagonal element, denoted $\sigma_j = \sum e_j^2$, is the mean square error for the jth sub-predictor. Assume that the matrix C can be inverted. Further, let A and B be two column vectors, $A = [\alpha_1(n), \ldots, \alpha_M(n)]^t$ and $B = \left[\frac{\beta}{2}, \frac{\beta}{2}, \ldots, \frac{\beta}{2}\right]^t$, where t denotes the transpose.

Then, it can be shown that the coefficients are given by

$$A = C^{-1}B. \tag{10}$$

3.2 Proposed algorithm

In our proposed algorithm, we make a further simplification by assuming that C is a diagonal matrix such that $\sigma_{jk} = \sum\sum e_j e_k = 0$ ($j \neq k$). Then it is easy to derive the coefficients from equations (8) and (9): $\alpha_j(n) = \frac{1}{D\sigma_j}$, where $D = \sum_{j=1}^{M} \frac{1}{\sigma_j}$ is a normalization factor. Therefore, the adaptive predictor for the current pixel (the nth pixel) is given by

$$P(n) = \frac{1}{D} \sum_{j=1}^{M} \frac{p_j(n)}{\sigma_j(n)}, \tag{11}$$

Here we change the notation of σ_j to $\sigma_j(n)$ to emphasize that it is a localized estimate of the mean square prediction error for the jth sub-predictor. To simplify the algorithm, instead of calculating $\sigma_j(n)$ over a block of pixels, a simple method to estimate $\sigma_j(n)$ is used. The estimate is given by the following equation:

$$\sigma_j(n) = \frac{1}{2}\left[\sigma_j(n-1) + R_j(n)\right] \tag{12}$$

where $R_j(n)$ is an estimate of the local mean square error for the jth sub-predictor. $R_j(n)$ is given by:

$$R_j(n) = e_j^2(n-1) + e_j^2(n-2) + e_j^2(n-3) + e_j^2(n-4) \tag{13}$$

where

$$e_j(n-m) = x(n-m) - p_j(n-m),\ m = 1,2,3,4$$

is the prediction error of the jth sub-predictor at location $n-m$. Equation (12) is a simple recursive filter that is employed to smooth out noise in the estimate and thus make the estimate more robust.

3.3 Discussion

There are two key elements in the APC scheme. One is the choice of sub-predictors, the other is the method used to determine the combination coefficients. In this section, we discuss issues related to these

$p_1 = W$
$p_2 = N$
$p_3 = N + W - NW$
$p_4 = NE$
$p_5 = (N + W)/2$
$p_6 = NW$
$p_7 = (NE + N)/2$

Table 1: List of fixed predictors. See Fig.1 for the location of the causal neighbouring pixels.

two elements and the relationship between the APC scheme and BMA.

1. The sub-predictors in Table 1 are all simple ones that only use those immediate neighbours of the current pixel. It is expected that the resulting predictor will perform well in smooth areas of an image where the pixels are strongly correlated to their nearest neighbours. However, this predictor may not perform well in areas having edges or textures where more pixels must be used to model the local statistical characteristics properly. A straightforward solution to this problem is to include a sub-predictor which uses more neighbouring pixels. Such a sub-predictor must be adapted to local characteristics of the image. The adaptation can be implemented by using the least-mean-square (LMS) algorithm or the least square (LS) approach by which the predictor coefficients are determined by minimizing the mean square prediction error of a block of causal neighbouring pixels. Compression results using the transform domain LMS algorithm based predictor and the LS-based predictor are reported in [7][35]. A more computationally demanding solution to this problem is to combine, instead of a group of simple fixed sub-predictors, a group of LS-based sub-predictors of different orders and different training window sizes [36]. Experimental results show that, as better predictors are used, the compression performance improves at the the costs of increased computational complexity.

2. The proposed combination scheme is also closely related to BMA. Referring to equation 3, the quantity θ_k can be interpreted as the prediction made by the kth sub-predictor given the data D. The probability $Pr(M_k|D)$ is proportional to $Pr(D|M_k)$ under the assumption that $Pr(M_k) = \frac{1}{K}$ is a

constant:

$$Pr(M_k|D) = \frac{Pr(D|M_k)Pr(M_k)}{\sum_{k=1}^{K} Pr(D|M_k)Pr(M_k)}$$

As a result, we have $\theta_B = \sum \theta_k Pr(M_k|D) \propto \sum \theta_k Pr(D|M_k)$. Therefore, the APC scheme described in the previous section can be regarded as a special case of BMA, where the combination coefficients are estimates of $Pr(D|M_k)$. The relationship between APC and BMA leads to several potential ways to improve the performance of APC. For example, since the probability $Pr(M_k)$ can be interpreted as our prior knowledge about the quality of the kth sub-predictor, it should be incorporated into the combination coefficient. It is also possible to use other methods to estimate the probability $Pr(D|M_k)$. Lee [13] has proposed a parametric estimation method where prediction errors are modeled as Laplace distributions.

4 Comparisons

The computational complexity of the APC is due to the calculation of the set of sub-predictors and the calculation of the combining coefficients. While the number of operations required for the latter is a function of the the number of sub-predictors, the number of operations required for the former can be very different for different sets of sub-predictors. Generally, the complexity of the APC is higher than predictors used in CALIC and LOCO which only require a few arithmetic and logical operations.

When comparing lossless image coding algorithms, we have to consider the compression performance as well as the computational complexity. The comparison task becomes more complicated if the fine tuning of the parameters of each algorithm is accounted for. In the following, we present a comparison of the compression performance.

Our experiments are performed on two sets of images. The first is the original 8 bits/pixel JPEG test image set. This set of images is widely used for comparisons in most image coding research. To test the effectiveness of our algorithm in medical applications, we include 8 medical images (12 bits/pixel) in our second set.

Table 2 shows the compression results of various lossless coding algorithms for the JPEG set. In this table, we use “APC-P7” to represent our proposed algorithm [6] which adaptively combines 7 fixed sub-predictors listed in Table 1. We also use “APC-LMS” to represent our proposed algorithm [7] which

	CALIC[33]	LOCO	HBB [26]	Lee[13]	APC-P7[6]	APC-LMS[7]	APC-LS[36]
baloon	2.78	2.90	2.80	2.79	2.75	2.72	2.65
barb2	4.46	4.69	4.48	4.47	4.46	4.39	4.28
barb	4.31	4.69	4.28	4.20	4.28	4.04	3.86
board	3.51	3.68	3.54	3.50	3.48	3.43	3.34
boats	3.78	3.93	3.80	3.76	3.75	3.70	3.60
girl	3.72	3.93	3.47	3.70	3.67	3.61	3.53
gold	4.35	4.48	4.37	4.35	4.34	4.30	4.25
hotel	4.18	4.38	4.27	4.24	4.19	4.15	4.10
zelda	3.69	3.89	3.72	3.68	3.68	3.63	3.54
Average bit rate	*3.86*	*4.06*	*3.89*	*3.85*	*3.85*	*3.77*	*3.68*

Table 2: Compression results (bits/pixel).

Image	(Size)	CALIC[26]	LOCO[26]	HBB[26]	APC-P6
c00156	(1755x1463)	5.82	5.93	5.71	5.72
chestxray	(1024x1024)	5.25	5.42	5.18	5.14
f005	(1799x1385)	6.25	6.35	6.16	6.14
100156	(2478x2478)	5.87	6.03	5.78	5.77
ref12b-0	(512x512)	2.77	2.90	2.68	2.62
ref12q5-0	(512x512)	2.83	2.95	2.66	2.60
skullpa-0	(512x512)	3.32	3.44	3.33	3.25
skullq7-0	(512x512)	3.33	3.40	3.23	3.20
Average bit rate		*4.43*	*4.55*	*4.34*	*4.31*

Table 3: Compression results (bits/pixel) of 12-bit medical images

adaptively combines the first 6 fixed predictors (in Table 1) with an *8*th order transform domain LMS-based adaptive predictor. "APC-LS" represents the APC scheme that combines a set of 7 linear predictors of different orders designed adaptively for each pixel by using the LS approach. In all of our experiments, a context based arithmetic coder is used to encode the prediction error. Of these three APC based algorithms, APC-P7 has the lowest level of computational complexity, APC-LMS is in the middle, and APC-LS has the highest level. Table 3 presents compression results for the medical image set, where the symbol "APC-P6" represents our proposed algorithm [6]. This algorithm adaptively combines the first 6 fixed sub-predictors shown in Table 1. Results shown in Tables 2 and 3 clearly show that the compression performance the APC scheme depends on the set of sub-predictors being combined. These results are also comparable to or better than those of the other compression algorithms.

5 CONCLUSIONS

In this paper, we have presented an overview of adaptive predictor combination for lossless image coding. We have discussed APC from a multi-disciplinary perspective and showed that it has been a useful tool for many scientific and engineering problems. In lossless image coding, APC has been demonstrated as a promising tool.

There are two essential elements in APC - the group of sub-predictors and the combination scheme. We have shown that by using a group of simple fixed sub-predictors, the compression performance is already comparable to, or better than, that of other published algorithms. We have also demonstrated that by using a group of LS based sub-predictors, the compression performance is greatly improved. Although our combination scheme described in this paper is based on the estimation of localized variance of prediction errors, it is closely related to Bayesian model averaging. Further improvement of the APC could be derived from the extensive theoretical and algorithmic study that has been carried out in fields such as forecasting and statistics.

References

[1] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Operations Research Quarterly*, 1969, v.20, pp. 319-325.

[2] R. G. Brown, "A new look at the Magill adaptive filter as a practical means of multiple hypothesis testing," *IEEE Trans. Circuits and Systems*, 1983, v. CAS-30, pp. 765–768

[3] H. G. Chen and A. Li, L. Kaufman and J. Hale, "A fast filtering algorithm for image enhancement," *IEEE Trans. Medical Imaging*, 1994, v. 13, pp. 557–564.

[4] Y. S. Chung and M. Kanefsky, "On 2-D recursive LMS algorithm using ARMA prediction for ADPCM encoding of images," *IEEE Trans. Image Processing*, 1992, v. 1, pp. 416–442.

[5] R. T. Clemen, "Combining forecasts: a review and annotated bibliography," *International Journal of Forcasting*, 1989, v.5, pp.559–584.

[6] G. Deng. and H. Ye, "Lossless image compression using adaptive predictor combination, symbol mapping and context filtering," *Proc. IEEE Int Conf. on Image Processing*, Kobe, Japan, 1999, v. 4, pp. 63–67

[7] G. Deng, "Adaptive prediction combination for lossless image coding," *Proc. European Signal Processing Conference*, CDROM, Tampere, Finland, 2000.

[8] P. Newbold and C. W. J. Granger, "Experience with forecasting univariate time series and the combination of forecasts," *Journal of the Royal Statistical Society A*, 1974, v.137 Part 2, pp.131–145.

[9] C. W. J. Granger, "Invited review: combining forecasts-twenty years later," *Journal of Forecasting*, 1989, v.8, pp.167–173.

[10] J. A. Hoeting, D. Madigan, A. E. Raftery and C. T. Volinsky, "Bayesian model averaging : A Tutorial," Technical Report 9814, Department of Statistics, Colorado State University, 1998.

[11] R. A. Jacobs, M. I. Jordan, S. J. Nowln and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, 1991, v.3, pp. 79-87.

[12] N. Kuroki, T. Nomura, M. Tomita and K. Hirano, "Lossless image compression by two-dimensional linear prediction with variable coefficients," *IEICE Trans. Fundamentals*, 1992, v. 7, pp. 882–889.

[13] W. S. Lee, "Edge adaptive prediction for lossless image coding," *Proc. Data Compression Conference*, Utah, USA, 1999, pp. 483–490.

[14] X. Li and M. T. Orchard, "Edge directed linear prediction for lossless compression of natrual images," *Proc. IEEE Int. Conf. Image Processing*, CDROM, Kobe, Japan, 1999

[15] D. T. Magill, "Optimal adaptive estimation of sampled stochastic process," *IEEE Trans. Automatic Control*, 1965, v. AC-10, pp. 434–439.

[16] G. Mealy and W. Tang, "Application of multiple model estimation to a recursive terrain height correlation system," *IEEE Trans. Automatic Control*, 1983, v. AC-28, pp. 323–331.

[17] N. D. Memon and K. Sayood, "Lossless image compression: A comparative study," *Still image compression, Proc. SPIE*, 1995, v. 2148, pp. 8-20.

[18] N. D. Memon, V. Sippy and X. Wu, "A comparison of prediction schemes proposed for a new lossless image compression standard," *Proc. IEEE International symposium on Circuits and Systems*, Atlanta, USA, 1996, v. 3, pp. 309–312.

[19] N. Memon and X. L. Wu, "Recent developments in context-based predictive techniques for lossless image compression," *The Computer Journal*, 1997, v. 40, pp. 127–136.

[20] B. Meyer and P. E. Tischer, "TMW-a new method for near lossless compression of greyscale images," *Proc. Picture Coding Symposium*, Berlin, Germany, 1997, pp. 533–538.

[21] B. Meyer and P. E. Tischer, "Extending TMW for lossless image compression," *Proc. Data Compression Conference*, Utah, USA, 1998, pp. 458–470.

[22] F. C. Palm and A. Zellner, "To combine or not to combine ? Issues of combining forecasts," *Journal of Forecasting*, 1992, v. 11, pp. 687–701.

[23] K. Popat and R. W. Picard, "Exaggerated consensus in lossless image compression," *Proc. IEEE International Conf. on Image Processing*, Austin, TX, USA, 1994, pp. 846–850.

[24] P. Roos, M. A. Viergever, M. C. Van Dijke and J. H. Peters, "Reversible intraframe compression of medical images," *IEEE Trans. Medical Imaging*, 1988, v. 4, pp. 328–336.

[25] T. Seemann and P. E. Tischer, "Generalized locally adaptive DPCM," *Technical Report No. 97/301*, Monash University, Australia, 1997.

[26] T. Seemann, P. E. Tischer and B. Meyer, "History-based blending of image sub-predictors," *Proc. Picture Coding Symposium*, Berlin, Germany,1997, pp. 147–151.

[27] A. E. Raftery, D. Madigan and J. A. Hoeting, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 1997, v. 92, pp. 179–191.

[28] P. T. Szymanski and D. Lemmon, "Adadaptive mixtures of local experts are source coding solutions," *IEEE International Conference on Neural Networks,* San Francisco, USA, 1993, v.3, pp.1391–1396.

[29] L. C. Wang, S. A. Rizvi and N. M. Nasrabadi, "A modular neural network vector predcitor for predictive VQ," *Proc. IEEE International Conf. on Image Processing*, Lausanne, Switzerland, 1996, v.3, pp. 443–447.

[30] X. Wang, "Optimal edge-preserving hybrid filters," *IEEE Trans. Image Processing*, 1994, v. 3, pp. 862–865.

[31] X. Wang, "On the gradient inverse weighted filter," *IEEE Trans. Signal Processing*, 1992, v. 40, pp. 482–484.

[32] M. J. Weinberger, G. Seroussi and G. Shapiro, "LOCO-I: A low complexity, context-based, lossless image compression algorithm," *Proc. Data Compression Conference*, Utah, USA, 1996, pp. 141–150.

[33] X. L. Wu, "An algorithmic study on lossless image compression," *Proc. Data Compression Conference*, Utah, USA, 1996, pp. 150–159.

[34] X. L. Wu and K. U. Barthel, "Piecewise 2D autoregrssion for predictive image coding," *IEEE Int. Conf. Image Processing*, Chicago, USA, 1998, pp. 901–904.

[35] H. Ye, G. Deng and J. C. Devlin, "Least squares approach for lossless image coding," *Proc. Fifth Internation Symposium on Signal Processing and Its Applications*, Brisbane, Australia, 1999, v.1, pp. 63–67.

[36] H. Ye, G. Deng and J. C. Devlin, "Least squares approach for lossless image coding," *Proc. IEEE Int Conf. on Image Processing*, CDROM, Vancouver, Canada, 2000.

[37] T. Zhang, "Compression by model combination," *Proc. Data Compression Conference*, Utah, USA, 1998, pp. 319–328.

List of Figures

List of Tables