A GENERAL FRAMEWORK FOR THE SECOND-LEVEL ADAPTIVE PREDICTION

Guang Deng and Hua Ye

Department of Electronic Engineering
La Trobe University, Bundoora, Victoria 3083, Australia
{d.deng, h.ye}@ee.latrobe.edu.au

ABSTRACT

In this paper we present a study of a general framework for second-level adaptive prediction which is formed from a group of predictors. It is a natural extension to that of the first-level which is formed directly from a group of pixels. The proposed framework offers a greater degree of freedom for adaptation and addresses some of the tough problems such as model uncertainty that is inherent to the first-level prediction methods. We show that the proposed methods of taking weighted average (WAVE) and weighted median (WMED) of a group of predictions are alternative and competitive adaptive image prediction methods. We have achieved better compression performance than that of TMW^{Lego} by combining a group of linear predictors.

1. INTRODUCTION

Adaptive prediction is a key component in lossless image coding. Adaptive prediction can be made either directly from a group of neighbouring pixels or from a group of component predictors. Therefore, we can regard adaptive prediction as having two levels. At the first level, the structure of the predictor is usually given and the parameters can be fixed or adaptively changed. The prediction is directly made from the pixels. There are well established theories [3] and algorithms, such as least squares (LS) [13] [5] and least mean square (LMS)[7], for first-level adaptive predictions. There are also some heuristics based methods to determine first-level predictors. For example, we can use one of a group of simple predictors shown in Table 1.

$p_1 = x(i, j-1)$
$p_2 = x(i-1, j)$
$p_3 = x(i-1, j) + x(i, j-1) - x(i-1, j-1)$
$p_4 = x(i-1, j+1)$
$p_5 = (x(i, j-1) + x(i-1, j))/2$
$p_6 = x(i-1, j-1)$
$p_7 = (x(i-1, j) + x(i-1, j+1))/2$

Table 1. List of fixed predictors. The current pixel is represented by $x(i, j)$ where *i* is the row index and *j* is the column index.

At the second level, we are given a group of N predictors, denoted by $\{p_n\}$ $(n = 1 : N)$. We would like to find out the most probable value for the current pixel *x*. In recent years, a number of methods that are related to the second-level prediction have been published. These methods employ the basic structures of either model selection or model combination. A representative method of model selection is the well known median adaptive prediction (MAP) used in the JPEG-LS standard [11]. The output of MAP is given by

$$x_{MAP} = MED\,(p_1, p_2, p_3) \tag{1}$$

which selects the median value of the three predictions (as shown in Table 1) as the final prediction. Another representative method is the gradient adjusted prediction (GAP) [12] which uses the local gradient and a set of thresholds to select a predictor from a group of simple predictors.

On the other hand, several researchers have proposed adaptive prediction techniques based on model combination. For example, Seemman [9] proposed combination scheme that penalizes "bad" predictors and the HBB (history-based blending) [10] algorithm which uses the LS approach to determine the combination coefficients. Deng [1, 2] proposed a combination scheme based on the variance of the predictor. Lee [4] proposed a predictor combination algorithm based on the Bayesian principle.

While there are abundant research results on the first-level of prediction, techniques based on the second-level have just emerged as powerful tools for lossless image compression problems. The motivation for our study is to develop a general framework for the second-level adaptive prediction such that it covers both model selection and model combination. We formulate the second level prediction problem as a parameter estimation problem. Based on the maximum likelihood principle, we propose two methods using weighted average and weighted median for model combination and model selection, respectively. These two methods are direct results of using a Gaussian and a Laplacian distribution in the optimization process. We carried out experiments using a group of simple predictors (shown in Table 1) to demonstrate the excellent predictive performance of the proposed second-level predictors. We also show that by combining a group of weighted-LS based linear predictors [14], compression performance better than that of TMW$^{\text{Lego}}$ [6] can be achieved. TMW$^{\text{Lego}}$ produced arguably the best (in terms of bit rate) lossless image compression results.

2. A GENERAL FRAMEWORK FOR THE SECOND-LEVEL PREDICTION

2.1. The main algorithm

We formulate the problem of the second-level prediction as the following:

$$\arg\max_{x} Pr(x|\{p_n\}, I)$$

where I represents the prior information. Using Bayes' rule, we have

$$Pr(x|\{p_n\}, I) \propto Pr(\{p_n\}|x, I) Pr(x|I) \quad (2)$$

where $Pr(x|I)$ is prior probability of *x*. To solve this optimization problem, we make two assumptions for the probability models (1) each prediction is statistically independent and (2) the prior distribution is flat for the whole range of possible values of *x*. A discussion of another setting for the prior is given in Section 2.2. We consider two types of models: Gaussian and Laplacian.

Using a Gaussian model, we have the probability density function (pdf)

$$Pr\,(p_n|x, I) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(x-p_n)^2}{2\sigma_n^2}} \quad (3)$$

where the mean is x and the variance is σ_n^2. The maximum likelihood (ML) solution to this problem is given by

$$x_G = \sigma_G^2 \sum_{n=1}^{N} \frac{p_n}{\sigma_n^2} \quad (4)$$

where $\frac{1}{\sigma_G^2} = \sum_{n=1}^{N} \frac{1}{\sigma_n^2}$. This is a weighted average of the component predictions.

Using a Laplacian model, we have

$$Pr\,(p_n|x, I) = \frac{1}{2\beta_n} e^{-\frac{|x-p_n|}{\beta_n}} \quad (5)$$

where as in the Gaussian model case, the quantity x is regarded as the mean and the variance is $2\beta_n^2$. The ML estimate for *x* is given by a weighted median filter output of the predictions using the respective $1/\beta_n$ as the weight. This is represented as

$$x_L = WM\,(\{p_n, 1/\beta_n\}_{n=1:N}) \quad (6)$$

The weighted median filter can be implemented in the following way. First the predictions are sorted such that

$$p_{(1)} \leq p_{(2)} \leq ... \leq p_{(N)} \quad (7)$$

where $p_{(n)}$ represents the *n*th largest prediction and its corresponding variance is represented by $\beta_{(n)}^2$. Then output of the weighted median filter is $x_L = p_{(L)}$ where L is the minimum index such that

$$\sum_{n=1}^{L} \frac{1}{\beta_{(n)}} \geq \frac{1}{2} \sum_{k=1}^{N} \frac{1}{\beta_{(k)}} \quad (8)$$

2.2. Prior probability and a sequential Bayesian solution

It is possible to set the prior to a distribution other than the flat one. For example, one possible setting is to regard the pdf of the random variable *x* as a Gaussian with the mean and variance being one of the prediction p_n and the corresponding σ_n^2, respectively. This is because it is reasonable to assume a particular model as the default model (prior) for an image. However, making such a setting does not change the solution to the problem.

For a particular choice of the prior, a sequential Bayesian solution can also be found. For example, if the first predictor in Table 1 is uses as the prior model, then its prediction value serves as the mean for the distribution, i.e., $Pr(x|I) \sim N(p_1, \sigma_1^2)$. The mean is also the minimum mean square error (MMSE) estimate of x (denoted by $y_1 = p_1$). Next, when another prediction (p_2) is known, the posterior probability is given by

	WAVE	WMED	MED	MIN	WMAP	MAP
baloon	2.87	2.92	3.12	3.05	3.12	3.12
barb2	4.94	4.98	5.26	5.02	5.12	5.18
barb	4.88	4.91	5.31	4.94	5.15	5.20
board	3.70	3.71	4.18	3.80	3.93	3.95
boats	4.13	4.16	4.54	4.22	4.29	4.31
girl	3.86	3.90	4.26	3.98	4.19	4.21
gold	4.62	4.64	4.91	4.75	4.74	4.72
hotel	4.53	4.56	4.94	4.66	4.73	4.73
zelda	3.80	3.86	4.01	3.97	4.08	4.11
AE	4.15	4.18	4.50	4.27	4.37	4.39

Table 2. The entropy (bits/pixel) of several prediction methods for the test images. AE represents the average entropy

$$Pr(x|p_2, I) \propto Pr(x|p_2) Pr(x|I) \quad (9)$$

The updated MMSE estimate is given by

$$y_2 = \frac{1}{D_2}\left(\frac{y_1}{\sigma_1^2} + \frac{p_2}{\sigma_2^2}\right) \quad (10)$$

where D_2 is defined as $\frac{1}{D_2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$. In the same way, when the (*N-1*)th prediction p_N is known, then the estimate is updated as:

$$y_{N-1} = \frac{1}{D_{N-1}}\left(\frac{y_{N-2}}{D_{N-2}} + \frac{p_N}{\sigma_N^2}\right) \quad (11)$$

where D_{N-1} is defined as: $\frac{1}{D_{N-1}} = \frac{1}{D_{N-2}} + \frac{1}{\sigma_N^2}$.

It can be shown that $y_{N-1} = x_G$. Comparing the ML solution (equation 4) to the sequential Bayesian solution (equation 11), we can see that the latter permits additional flexibility that we can stop the calculation at an index *k* ($k < N$)and use y_k as the final output.

2.3. Estimation of the variance

A vital task in the proposed framework is to estimate the variance of each prediction. Since images are usually regarded as nonstationary signals, the estimation must be robust and localized for each pixel. We have proposed a simple algorithm for this purpose [2]. We assume that the current pixel is the *m*th pixel at location index (i, j) of the image. We denote the estimate of the variance for the nth predictor for the current pixel as $\sigma_n^2(m)$, which is calculated by the following equation:

$$\sigma_n^2(m) = \frac{1}{2}\left[\sigma_n^2(m-1) + \frac{1}{4}E_n(m)\right] \quad (12)$$

where $\sigma_n^2(m-1)$ is the estimate of the variance of the previous pixel for the *n*th predictor, and $E_n(m)$ is an estimate of the "local" mean square error for the *n*th predictor. Let $e_n\,(i, j)$ represent the prediction error of the *nth* predictor for the pixel at location (i, j). $E_n(m)$ is given by

$$E_n(m) = e_n^2(i, j-1) + e_n^2(i-1, j-1) + e_n^2(i-1, j) + e_n^2(i-1, j+1). \quad (13)$$

Equation (12) is a simple recursive filter to smooth out noise in the estimate.

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we have performed experiments using a set of JPEG test images which are 576×720, 8 bits/pixel. The proposed adaptive prediction methods using weighted average and weighted median are represented by WAVE and WMED, and are defined by equations 4 and 6, respectively. The method denoted by MED is a special case of WMED where the weights are regarded as equal. The method denoted by MIN selects the prediction with the minimum variance as the final prediction. An extension to MAP, denoted by WMAP, is given by

$$x_{WMAP} = WM\left(\{p_n, \beta_n\}_{n=1:3}\right) \quad (14)$$

The performance of each method is measured by the entropy of the prediction errors of each image. For comparison reason, we use MAP as a benchmark. The results are listed in Table 2. From this table, we can see that the performance of the proposed adaptive prediction methods (WAVE and WMED) is consistently better than that of MAP for all test images. The average entropy is improved by more than 0.2 bit/pixel. The MIN method is actually a performance based prediction selection method that selects the best performing predictor output as the final prediction. It also outperforms MAP, although its performance is a bit inferior to those of WAVE and WMED. It is also interesting to note that extending MAP to WMAP results in slight improvement in performance.

The MED predictor can be regarded as a direct extension to MAP by having more predictions. However, this simple extension does not work well. This can be explained by the proposed general framework. Taking the median value as the final prediction is an optimal choice only if all of the predictions follow the same Laplacian model (with same mean and variance). However, for a group of predictors such as those defined in Table 1, different predictors have different variances for an area of the image. Thus, with the Laplacian model, the weighted median is the proper method for prediction selection and is a natural extension to MAP.

In a lossless image coding algorithm, the prediction error is coded by using an entropy coding algorithm. Context-based adaptive coding is aimed at capturing the remaining redundancy and nonlinear relationship among prediction errors. In other words, the relatively poor performance of a predictor can be compensated by context-based adaptive coding. Therefore, it is necessary to compare the performance of different prediction methods when they are used with a context-based coding method.

To perform such a comparison, we used a context-based arithmetic coder [1] to encode the prediction errors. The bit rate for each image is shown in Table 3. The difference between the average entropy of the prediction errors and the average bit rate of the entropy coding results for each prediction method is shown in Table 4. From these two tables we can see that while the entropy coding method does help improve the compression performance for each prediction method, the relatively poorly performing prediction methods gain more. However, the ranking of the prediction methods remains the same. A better performing prediction method results in a lower bit rate. For example, the average bit rate for WAVE is about 0.1 bit/pixel better than that of MAP. These results highlight the importance of using a good predictor in a lossless image coding system.

Although there is no restriction as to what type of prediction method should be included in the proposed adaptive prediction method, we have selected a group of relatively simple predictors to keep the computational complexity comparable to that of MAP.

	WAVE	WMED	MED	MIN	WMAP	MAP
baloon	2.75	2.78	2.84	2.82	2.84	2.84
barb2	4.45	4.51	4.60	4.51	4.55	4.58
barb	4.27	4.32	4.50	4.34	4.40	4.46
board	3.49	3.50	3.65	3.52	3.55	3.57
boats	3.76	4.79	3.92	3.81	3.83	3.85
girl	3.67	3.71	3.86	3.73	3.82	3.84
gold	4.33	4.35	4.45	4.38	4.38	4.39
hotel	4.19	4.21	4.35	4.23	4.26	4.27
zelda	3.67	3.71	3.78	3.73	3.77	3.80
ABR	3.84	3.88	3.99	3.90	3.93	3.95

Table 3. The entropy coding results (bits/pixel) of several prediction methods for the test images. ABR represents the average bit rate

WAVE	WMED	MED	MIN	WMAP	MAP
0.31	0.30	0.51	0.37	0.44	0.44

Table 4. The difference between the average entropy of the prediction errors and the average bit rate of the coding results for each prediction method.

	TMW$^{\text{Lego}}$	WAVE-WLS
baloon	2.60	2.60
barb2	3.84	3.75
barb	4.24	4.18
board	3.27	3.27
boats	3.53	3.53
girl	3.47	3.45
gold	4.22	4.20
hotel	4.01	4.01
zelda	3.50	3.51
Average bit rate	3.63	3.61

Table 5. The compression results (bits/pixel) using a combination of weighted-LS based linear predictors and TMW$^{\text{Lego}}$ algorithm .

It should be noted that if computational complexity is not a major concern, then it is possible to improve the predictive performance of the proposed method by using better predictors. The limited performance of predictors listed in Table 1 is mainly because only four causal neighbouring pixels are used. Therefore, a more powerful prediction should involve more pixels. For example, by combining a group of weighted-LS based linear predictors [14], we have demonstrated algorithms with compression performance better than that of TMW$^{\text{Lego}}$ which has achieved arguably the best compression performance. The comparison results are shown in Table 5.

At the end of this section, we discuss two reasons for the superior performance of the second-level prediction. One reason is its greater ability for adaptation. Methods of the first-level prediction usually assumes a fixed model structure. For a linear predictor, it is difficult to adaptively change both the order and the coefficients at the same time. Model averaging provides a mechanism to solve this problem. For example, one can combine a group of linear predictors of different orders. These predictors can be either fixed or adaptive. A combination of these predictors results in a predictor that can adjust its structure and parameters simultaneously.

The other reason is that WAVE provides a mechanism to tackle the problem of model uncertainty that is inherent to almost all first-level predictors. For example, how do we determine and justify the order of a linear model for an image? In an extreme case, there may be a different optimal order for each pixel. In addition to this problem, when using the least squares (LS) method to design a linear predictor for each pixel the size and the shape of the training window are also uncertain [13] [5]. These problems are seldom addressed in the first-level prediction.

Model uncertainty can be expressed as the posterior probability of the model being the correct model for the current pixel, or more generally, as an estimate of the predictive performance of the predictor for the current pixel. According to the theory of Bayesian model averaging[8], given the data *D* and a group of *N* models M_n, the posterior distribution is determined by:

$$P(x|D) = \sum_{n=1}^{N} P(x|D, M_n)P(M_n|D) \tag{15}$$

The MMSE estimate of *x* is given by

$$y = E\left[x|D\right] = \sum_{n=1}^{N} y_n P(M_n|D) \tag{16}$$

where $y_n = E\left[x|D, M_n\right]$ is MMSE estimate of *x* using the *n*th model and can be regarded as the prediction output for the *n*th model. The quantity $P(M_n|D)$ is the probability that the *n*th model makes a correct prediction. This probability can be regarded as a measure of the degree of belief that the *n*th model is the true model that generates the sample x. One way to express our degree of belief is to use a localized estimate of the variance of the prediction errors of the *n*th model and let

$$Pr(M_n|D) \propto \frac{1}{\sigma_n^2} \tag{17}$$

Therefore, the combination coefficient can be regarded as a representation of the uncertainty of each model. By averaging over a group of models, the uncertainty is reduced and the predictive performance is improved.

4. CONCLUSION

In this paper we have presented a general framework for the second-level adaptive prediction which is based on the principle of maximum likelihood. The second-level prediction is a natural extension to that of the first-level. It offers a greater degree of freedom for adaptation and addresses some of the tough problems such as model uncertainty that is inherent to the first-level prediction methods.

The proposed methods of taking weighted average (WAVE) and weighted median (WMED) of a group of predictions are alternative and competitive adaptive image prediction methods. Compared to some published methods which are designed based on heuristics and have some parameters that require fine-tuning for particular images, the proposed methods are based on well developed principles. In addition, the proposed WAVE and WMED methods permit a trade-off between the predictive performance and the computational complexity. This is accomplished by carefully selecting a group of predictors. We have shown that by using a group of weighted-LS based predictors, better compression results than those of TMW^{Lego} can be obtained.

5. REFERENCES

[1] G. Deng, “Transform domain LMS-based adaptive prediction for lossless image coding,” *Signal Processing Image Communications*, vol. 17, pp. 219–229, 2002.

[2] G. Deng, H. Ye, and L. W. Cahill, “Adaptive combination of linear predictors for lossless image compression,” *IEE Proc.-Sci. Meas. Technol.*, vol. 147, pp. 414–419, Nov. 2000.

[3] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, New Jersey, USA: Prentice Hall Inc., third ed., 1996.

[4] W. S. Lee, “Edge adaptive prediction for lossless image coding,” in *Proc. IEEE Data Compression Conference*, pp. 483–490, Snowbird, Utah, Mar 1999.

[5] X. Li and M. Orchard, “Edge-directed prediction for lossless compression of natural images,” *IEEE Trans. Image Processing*, vol. 10, pp. 813–817, June 2001.

[6] B. Meyer and P. Tischer, “TMW^{Lego} – an object oriented image modelling framework,” in *Proc. IEEE Data Compression Conference*, Snowbird, Utah, Mar 2001.

[7] G. Qiu, “A progressively predictive image pyramid for efficient lossless coding,” *IEEE Trans. Image Processing*, vol. 8, pp. 109–115, Jan. 1999.

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

[9] T. Seemann and P. E. Tischer, “Generalized locally adaptive DPCM,” Technical Report 97/301, Department of Computer Science, Monash University, 1997.

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

[11] M. J. Weinberger, G. Seroussi, and G. Sapiro, “The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS,” *IEEE Trans. Image Processing*, vol. 9, pp. 1309–1324, Aug. 2000.

[12] X. Wu and N. D. Memon, “Context-based, adaptive, lossless image coding,” *IEEE Trans. Commun.*, vol. 45, pp. 437–444, Apr. 1997.

[13] H. Ye, G. Deng, and J. C. Devlin, “Adaptive linear prediction for lossless coding of greyscale images,” in *Proc. IEEE International Conference on Image Processing*, Vancouver, Canada, 2000.

[14] H. Ye, *A study on lossless compression of greyscale images*. PhD thesis, Department of Electronic Engineering, La Trobe University, Oct. 2002.