Adaptive interpolator with context modeling in lifting scheme for lossless coding

Wen-Thong Chang & Wen-Jen Ho,
Department of Communications Engineering, National Chiao Tung University,
1001 University Ave, Shinchu, Taiwan

ABSTRACT

Adaptive interpolation with double-interpolator is used for image coding. The double-interpolator means that two adaptive prediction stages are used. The outer loop is an adaptive FIR predictor. The inner loop is an texture based bias estimation. The bias means a content dependent estimated prediction error. By assuming the signals to consist of polynomials of various degrees, the predictor of the outer loop is constructed by linearly combining a set of maximally flat filters. The maximally flat filter is the filtering implementation of the Lagrange interpolation. The prediction is done block by block. Within a block, similar with the lifting scheme, a hierarchical multiresolution prediction is used starting from the lowest resolution 2 by 2 sub-block. The least square prediction error criterion is used to derive the weighting coefficients of the predictor. To further reduce the prediction error, an inner loop to estimate the prediction error is included. Within the inner loop, the pixel to be predicted is classified into groups according to the neighborhood condition. The accumulated mean of the prediction errors of all the pixels within the same group is considered as the bias of the prediction error. This bias is then extracted from the actual prediction error to reduce entropy.

Keywords: Adaptive prediction, lifting scheme, multi-resolution prediction, image coding.

1. INTRODUCTION

Predictive coding has been the major approach for lossless coding. For coding efficiency, several issues are considered in this paper. In order to reduce the prediction error, a double-predictor is used. The outer-loop is a conventional predictor. While, a second inner loop is used to model the quantity of the prediction error. The difference of the true prediction error and the estimated prediction error is encoded. The second issue is the design of the predictor. By assuming a polynomial signal model, Lagrange interpolator is used. To account for the variation of the signal, an adaptive interpolator with the linear combination of a set of Lagrange interpolator is used. For multi-resolution purpose, a multi-rate prediction scheme is used. This signal representation scheme is similar to that used in the lifting scheme.

In lifting, a predictor is applied to the even (or odd) samples to predict the odd (or even) samples and the interpolated results are subtracted from the original odd (or even) samples. The lifting prediction has been shown to be equivalent to filter bank decomposition. However, in each lifting step, the predictor output is downward truncated for an integer to integer transform. There are usually two prediction steps associated with a lifting scheme. The dual lifting step can be regarded as a prediction stage to predict the highpass components and the primal lifting step can be regarded as an update stage to update the lowpass components. In our approach, the primal lifting step used to update the lowpass components is kept fixed. In the extreme case, if the primal lifting is omitted, the analysis lowpass filter is a delta function and the synthesis lowpass

In *Visual Communications and Image Processing 2000*, King N. Ngan, Thomas Sikora, Ming-Ting Sun, Editors, Proceedings of SPIE Vol. 4067 (2000) • 0277-786X/00/$15.00

filter is a halfband filter.

By assuming a polynomial signal model, the optimal interpolator is then the Lagrange interpolator. The tap length and the zero distributions of the filter are related to the order of the polynomials that are to be interpolated. This is the so-called A_p condition[1]. An immediate consequence of this theorem is that the highpass components of the filter bank will be zero if the signals to be processed are polynomial of degree N-1 and the filter has N zeros at π. In this particular case, the optimal predictor to minimize the highpass components of the filter bank, or the prediction error, is the maximally flat (Maxflat) filter. The maximally flat filter is the filter implementation of the Lagrange interpolation. For practical situations, adaptive filter based on the linear combination of a set of maximally flat filters is used. The weightings of the combination can reflect the image properties in the original pixel domain.

There is another set of bases that can be used to represent the maximally flat filter. This is the Bernstein basis. With the Bernstein basis, some image properties in the frequency domain can be obtained. Thus the predictor can also be represented as a linear combination of the Bernstein filters[3]. These two representations can be related by a change of the bases that are used to represent the predictor. The Bernstein filter is a bandpass filter. Thus the weightings can indicate the image properties in the frequency domain. Actually, any filter can be represented as a linear combination of a set of max-flat filters or Bernstein filters. These weighting variables can be used as content index in the subsequent inner loop for prediction error modeling.

The function of the inner loop is to reduce further the prediction error[4]. This technique has been applied in the content modeling with sequential prediction[2]. The application of this technique in the multirate case[5] is more complicated with the hierarchical structure, because the distributions of the prediction error in each layer are different. The amount of the prediction error is estimated based on the neighboring information of the pixel that is currently under prediction. Usually, prediction errors from pixels with similar pattern are collected and processed to form an estimate of the prediction error for that class.

2. FILTER DESIGN

The prediction is a feed-forward type. The design of the predictor to be used in this scheme is considered in this section. We focus on the dual step with the constraint to minimize the highpass energy based on the known lowpass components. The implementation of the integer dual lifting scheme is illustrated in Fig. 1. Formally, a top-down decomposition procedure can be described as, for $j = 1 \sim J$,

$$\begin{aligned} c^j[k] &= c^{j-1}[2k], \\ \bar{d}^j[k] &= d^{j-1}[2k+1], \\ &= c^{j-1}[2k+1] - \left\lfloor \sum_n c^j[n] p[2n-(2k+1)] \right\rfloor, \\ &= c^{j-1}[2k+1] - \left\lfloor \sum_n c^j[n] e^1[n-k-1] \right\rfloor, \end{aligned} \tag{1}$$

where $e^1[k] = p[2k+1]$ is the 1st polyphase component of $p[k]$. For reconstruction, $c^j[k]$ can be obtained with prediction from its lower resolution versions plus the prediction errors. A bottom-up reconstruction procedure is, for $j = 1 \sim J$,

$$\begin{aligned} c^{j-1}[2k] &= c^j[k], \\ c^{j-1}[2k+1] &= \bar{d}^j[k] + \left\lfloor \sum_n c^j[n] p[2n-(2k+1)] \right\rfloor, \\ &= \bar{d}^j[k] + \left\lfloor \sum_n c^j[n] e^1[n-k-1] \right\rfloor. \end{aligned} \tag{2}$$

In this multi-rate case, if the truncation is neglected, an associated set of biorthogonal bases exists. The scaling function is called the interpolative scaling function. That is $\Phi(t)$ is cardinal, i.e. $\Phi(k) = \delta(k), k \in Z$. According to the two-scale relation,

$$\Phi(t) = \sum_k p[k] \Phi(2t-k) \tag{3}$$

the values of $\Phi(t)$ on the half-integer points can be computed by $\Phi\left(\frac{m}{2}\right) = \sum_k p[k]\Phi(m-k)$.

Consider a continuous function $f(t)$ with derivative $f^{(i)}(t)$, $0 \le i \le N$. It can be shown that if $\Phi(t)$ satisfies the condition,

$$\sum_{k=-\infty}^{\infty} (t-k)^m \Phi(t-k) = \delta(m), m = 0 \sim N-1. \tag{4}$$

then $\left\| I_j f(t) - f(t) \right\|_c \le \frac{M 2^{jN}}{N!} \left\| f^{(N)}(t) \right\|_c$, where M is a constant and the norm $\left\| f(\cdot) \right\|_c = \sup_{t \in R} |f(t)|$. The term $\frac{M 2^{jN}}{N!} \left\| f^{(N)}(t) \right\|_c$ is the upper bound of the approximation error when the basis $\Phi(t)$ satisfying Equation (4) is used.

Now let us concentrate on some basic properties of the basis. With the two-scale relation, the property of the basis shown in equation (4) can be related to the property of the filter $p[k]$. Thus the accuracy of the interpolation can be determined from the property of the filter. The highest degree of the interpolated polynomial can be determined from $p[k]$ or from $p(e^{jw})$.

Thus if $\hat{\Phi}(\omega)$ satisfies the Strang-Fix condition, the corresponding requirement on the filter $p[k]$ is that it has N zeros at π, i.e. the transfer function $p[e^{jw}]$ can be represented as

$$p(e^{j\omega}) = (1+e^{j\omega})^N R(e^{j\omega})$$

where $R(e^{j\pi}) \neq 0$. A filter with N zeros at $\omega = \pi$ is a N-regular filter. With the N -regular filter, the corresponding scaling function and its translates reproduce polynomials up to degree $N-1$. From the time-domain Strang-Fix condition, this fact can also be seen. The condition in equation (4) can be rewritten in the following discrete form as,

$$\sum_k (\frac{m}{2} - k)^n p[m-2k] = \delta[n], n = 0 \sim N-1,$$

if $t = \frac{m}{2}$ is considered. This property is the sum rule : $\sum_k (-1)^k k^n p[k] = 0, n = 0 \sim N-1,$ of the halfband filter. This is the A_p condition indicating that there are N zeros at $\omega = \pi$.

The above mentioned Strang-Fix condition indicates that an N-regular halfband filter can be used to interpolate a polynomial of degree $N-1$. The minimum length of an N-regular halfband filter is 2N-1 and is called the maximally flat filter. The above derivation shows the results that can be obtained when the maximally flat filter is used for interpolation. The above procedure indicates how the sampled points from $f(t)$ can be used to interpolate the same polynomial function $f(t)$.

If the practical signal is treated as a combination of various-order polynomials, the predictor is to be composed of many maxflat halfband filters. In the beginning of this section, we have mentioned the design of a non-Maxflat filter by linear combination of a set of Maxflat filters. The polyphase matrix consisting of the first polyphase components of a set of Maxflat filters with different lengths is defined as

$$E_m = \begin{bmatrix} e_N[-N] & 0 & 0 & 0 \\ e_N[-N+1] & e_{N-1}[-N+1] & 0 & 0 \\ \vdots & \vdots & \cdots & e_1[-1] \\ \vdots & \vdots & \cdots & e_1[0] \\ e_N[N-2] & e_{N-1}[N-2] & 0 & 0 \\ e_N[N-1] & 0 & 0 & 0 \end{bmatrix}$$

where E_m^{half} denotes the $N \times N$ upper-half matrix of E_m and $\tilde{E}_m^{half}$ denotes the $N \times N$ lower-half matrix of E_m. So, the design of a non-Maxflat filter can be described as

$$e^1 = e_N + E_m \bar{\beta},$$

where $\bar{\beta} = [\beta_N, \beta_{N-1}, \cdots, \beta_1]^T$ is the weighting vector of the combination.

3. SIMULATION RESULTS

The context modeling consists of two techniques in this implementation. The first is the way how the entropy coding is used to encode the prediction error. In this implementation, the set partitioning is used. The second issue is the bias cancellation in the prediction. The bias cancellation means that for each pixel an extra prediction is used to predict the prediction error that will be generated with the adaptive predictor. This bias is then extracted from the generated prediction error to further reduce the quantity of the prediction error that is to be encoded. If the bias is a constant, it can be simply included in the predictor. The real situation is that it is usually content dependent. Although an adaptive predictor has been used in our implementation to address the content dependence in the prediction. The adaptation algorithm to adjust the predictor coefficient is still very local and the texture relation among the neighboring pixels is still not considered in order to limit the complexity of the adaptation algorithm and the prediction process. Thus, an adaptive bias cancellation is considered. The bias is considered to be texture dependent to complement the task of the adaptive predictor.

Thus, based on the texture information, a set of groups is constructed. For each group, a bias is estimated. This bias is the accumulated mean of the prediction error of all the pixels in the group.

In our simulations, the adaptive prediction is the outer loop and the bias estimation is the inner loop. The adaptive lifting scheme performs better than the non-adaptive S and S+P transform and is comparable with the LOCO[7]-1 with the median edge detector (MED) and the CALIC[2] with the gradient adjusted predictor[6] (GAP). However, LOCO-I and CALIC are single-resolution predictive coding scheme without progressive decoding capability. We focus on the dual step with the constraint to minimize the highpass energy based on the known lowpass components.

To obtain the prediction errors, the interpolation process is executed hierarchically with a pyramid structure. The block size is 32*32 corresponding to four decomposition levels. The resultant representation for each block consists of the prediction errors hierarchically from 2*2 to 4*4, from 4*4 to 8*8, etc. The results with only the outer loop are listed in Table 1. The prediction error is encoded with the set partitioning algorithm. Within the inner loop, the contexts to model the prediction errors consist of the texture contexts and the energy contexts. The prediction errors with the similar context are grouped together and the mean of the prediction errors is used as the output of the inner loop. In CALIC, the GAP prediction error is compared with this content information to generate an improved prediction error. For the multirate case, the contexts or neighborhood conditions depend on the pyramid levels.

The texture information is the pixel distribution of the neighboring pixels. To form the texture information, the values of these pixels that are used with the predictor are compared with the prediction output. If the pixel value is larger than the prediction output, it is labeled as 1. Otherwise, it is labeled as 0. This indicates the relations among the pixels that are used to predict the current pixel. Thus, for each pixel, the neighborhood information will consist of a sequence of binary digits. These binary sequences are used as a part of the texture information. In multirate prediction, the prediction is done from lower resolution toward higher resolution, the neighborhood characteristics are different in each level. Thus, the content information in each level is formed separately. Since the neighborhood relation is tighter in the upper level, the prediction errors decrease as prediction goes from the lower resolution toward the higher resolution. The necessity to separately form content information in each level is a drawback of this multirate prediction. Because the sample size in each level is made smaller such that the variation of the content mean is larger than that of the sequential case. This is the reason why the contribution of the inner loop in the multirate case is not as good as that in the sequential case.

Let us denote these pixels that are used for prediction as x_k. The prediction is then $\hat{x} = \sum_{k=1}^{k} p_k x_k$. Let t_k denote the state of pixel x_k. Then t_k=1, if $x_k > \hat{x}$, otherwise t_k=0. The sequence t_k is the content information. The energy can be defined as $\varepsilon = \sum w_k (x_k - \hat{x})^2$. The error amount is also quantized into L levels and denoted with $\log_2 L$ bits. The total context information is the combination of the t_k sequence and the $\log_2 L$ bits. The quantized contexts are denoted as $C(\varepsilon, t_k)$. The inner loop is to estimate $\bar{e}(\varepsilon, t_k) = E[e|c(\varepsilon, t_k)]$, with e= (x-$\hat{x}$). The final prediction is then $\tilde{x} = \hat{x} + \bar{e}(\varepsilon, t_k)$. The actual transmitted error is x-$\tilde{x}$.

Another consideration of the content information is the energy for each pixel. The energy content is the sum

of prediction errors of some related pixels. The spatial relation of pixels for this energy context is shown in Fig.2. For the LH (low-high) band and the HL (high-low) band, the energy is formed with the four neighboring prediction errors and the error of the parent pixel. For the HH (high-high) band, the two corresponding coefficients in the LH band and HL band are used. Prediction errors within the context are then quantized into five levels. Energy contexts are formed by quantizing the error energy in LH, HL and HH bands, respectively, as

$$\Delta_{LH} = \delta + a_{lh}|nn| + b_{lh}|p_n|;$$
$$\Delta_{HL} = \delta + a_{hl}|ww| + b_{hl}|p_w|;$$
$$\Delta_{HH} = \delta + a_{hh}|c_{lh}| + b_{hh}|c_{hl}|$$

where p_n and p_w are the parent prediction errors to the north and west, c_{lh} and c_{hl} are the brother prediction errors of the current pixels. δ^5 represents the amount from the pixels that are all considered in all the three bands. We chose a_{lh}, a_{hl}, a_{hh}, b_{lh}, b_{hl} and b_{hh} all to be 1. Energy contexts are quantized to 8 levels. Then, the error modeling is done by using 5^5 texture contexts and 8 energy contexts. A total of 25000 compound contexts. With the context-based error modeling, the entropy of the prediction errors for different images is listed in the third column in Table 2. The improvement is about 1.62% from 4.33 bpp to 4.26 bpp.

There can be many reasons that the improvement in the multirate case is less than that in the sequential case. Several possible reasons are 1) the setting of the threshold values, 2) the selection of context information, 3) the neighborhood relationship is weaker in the multirate environment (especially in the lower resolution levels), etc. The error modeling has no effect in the lower resolution levels. Since most of the large prediction errors are found in these levels, how to utilize the error modeling at these levels becomes an interesting problem. The entropy improvement of the prediction errors at different resolution levels are shown in Table 3. It can be seen that the entropy reduction by the error modeling is only significant at the fine resolution level. The coding results for these various images are shown in Table 4. To encode the prediction error, the set partitioning in hierarchical trees (SPIHT) method is used.

REFERENCES

1. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.
2. X. Wu, "Lossless compression of continuous-tone images via context selection, quantization, and modeling", IEEE Trans. IP, vol. 6 no.5 , pp.656-664.1997
3. H. Caglar, A. N. Akansu "A generalized parametric PR-QMF design technique based on Bernstein polynomial approximation" IEEE Trans. SP. pp.2314-2321 1993.
4. D. G. Daut, D. Zhao, J. C. Wu, "Double predictor differential pulse code modulation algorithm for image data compression" Optical Engineering, vol. 32, pp. 1514-1523, July 1993.
5. N. Memon, X. Wu and B-L Yeo, "Improved techniques for lossless image compression with reversible integer wavelet transforms", IEEE ICIP, pp.891-895. 1998.
6. N. Memon, V. Sippy and X. Wu, "A comparison of prediction scheme proposed for a new lossless image compression standard", IEEE ISCAS, pp. 309-312. 1996
7. S.A. Martucci, "Reversible compression of HDTV images using median adaptive prediction and arithmetic coding", IEEE ISCAS, pp.1301-1313. 1990

Table 1: ENTROPY OF PREDICTION ERRORS OF TEST IMAGES

Images	MED	GAP	Adaptive IDPCM pyramid
couple	2.74	3.00	2.80
face	4.79	4.65	4.61
girl	4.17	4.08	4.09
hat	4.50	4.41	4.40
jet	3.93	4.25	4.28
lena	4.55	4.39	4.35
london	3.78	3.93	3.77
baboon	5.77	6.26	6.04
MRI1	4.44	4.40	4.24
MRI2	4.74	4.68	4.58
MRI3	4.40	4.34	4.29
MRI4	4.58	4.54	4.53
MRI5	4.45	4.41	4.37
Average	4.37	4.41	4.33

Table 2: ENTROPY OF PREDICTION ERRORS OF TEST IMAGES

Images	CALIC	Adaptive IDPCM pyramid + error modeling
couple	2.57	2.79
face	4.58	4.59
girl	3.97	4.08
hat	4.31	4.18
jet	4.17	4.02
lena	4.29	4.25
london	3.74	3.75
baboon	6.16	6.09
MRI1	4.17	4.24
MRI2	4.47	4.43
MRI3	4.15	4.18
MRI4	4.34	4.51
MRI5	4.21	4.32
Average	4.24	4.26

Table 3: AVERAGE ENTROPY AT DIFFERENT RESOLUTION LEVELS

	$\bar{d}^1[k]$	$\bar{d}^2[k]$	$\bar{d}^3[k]$	$\bar{d}^4[k]$
No error modeling	4.18	4.62	5.38	6.00
Have error modeling	4.10	4.61	5.38	6.00

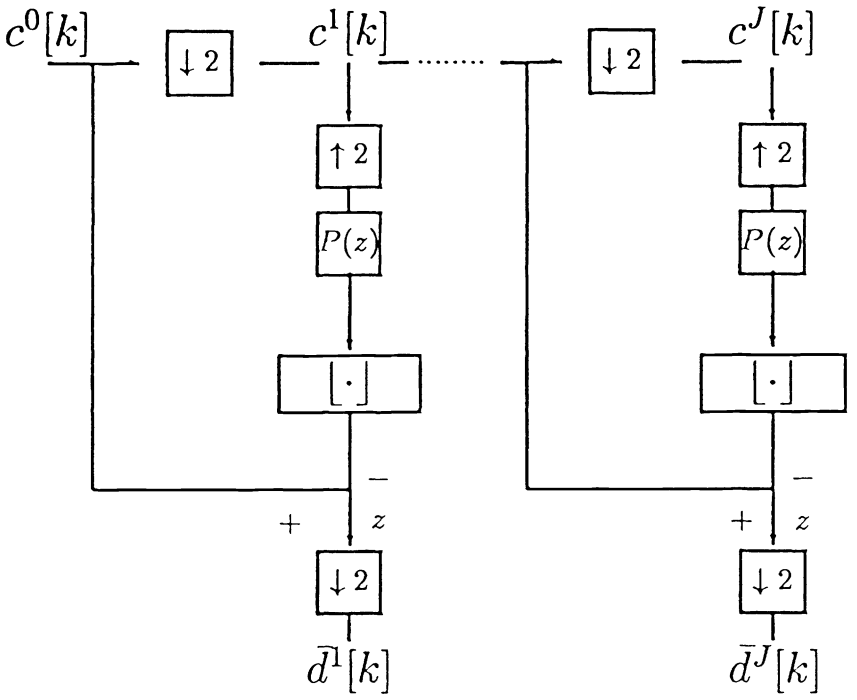

Figure 1: THE DECOMPOSITION STRUCTURES OF THE INTERPOLATIVE DPCM PYRAMID.

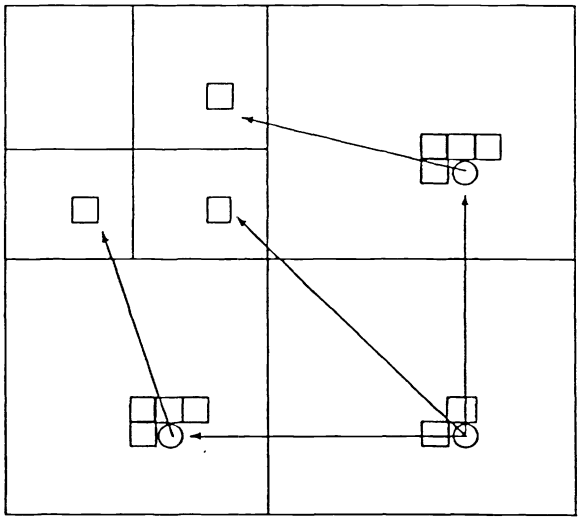

Figure II: CONTEXTS USED FOR TEXTURE CONTEXTS