Lossless Image Compression Based on Optimal Prediction, Adaptive Lifting, and Conditional Arithmetic Coding

Nikolaos V. Boulgouris, *Student Member, IEEE*, Dimitrios Tzovaras, and Michael Gerassimos Strintzis, *Senior Member, IEEE*

***Abstract*—The optimal predictors of a lifting scheme in the general n-dimensional case are obtained and applied for the lossless compression of still images using first quincunx sampling and then simple row–column sampling. In each case, the efficiency of the linear predictors is enhanced nonlinearly. Directional postprocessing is used in the quincunx case, and adaptive-length postprocessing in the row–column case. Both methods are seen to perform well. The resulting nonlinear interpolation schemes achieve extremely efficient image decorrelation. We further investigate context modeling and adaptive arithmetic coding of wavelet coefficients in a lossless compression framework. Special attention is given to the modeling contexts and the adaptation of the arithmetic coder to the actual data. Experimental evaluation shows that the best of the resulting coders produces better results than other known algorithms for multiresolution-based lossless image coding.**

***Index Terms*—Arithmetic codes, image coding, wavelet transforms.**

I. Introduction

WAVELET decomposition has established itself as one of the state of the art techniques for image coding problems because of its capability for allowing the generation of lossy versions of an original image at multiple resolutions and bitrates. Many applications such as the transmission of depth maps for the construction of 3-D views of a scene [2] or the efficient storage and communications of medical images require lossless coding [3], [4]. Most known lossless coders are based on predictive decorrelation [5], [6] and do not have any preview capability. In this paper, we propose lossless coders based on wavelet transform that perform within a few percentage points in comparison to state-of-the-art context-based coders (which do not have preview capability, i.e., progressive transmission).

A wavelet transform is realized using filter banks which split the image information into frequency subbands. Due to their inherent property of producing floating point output, classical filter banks can not in general be used in lossless compression schemes, since the coding cost for the coding of the floating point wavelet coefficients would be prohibitively large. Instead, interpolative pyramids may be used [7]–[9].

The lifting scheme has recently attracted much interest. It is a way to implement critically sampled filter banks which have integer output. The fundamentals of lifting can be found in [10]. An algorithm for decomposing wavelet transforms into lifting steps was described in [11]. A complete presentation and evaluation of the lifting scheme for the separable case was given in [12]. Extension to the general multidimensional case was presented in [13]. Finally, nonlinear wavelet transforms were recently investigated in [14]–[16].

Following decorrelation of the input image by methods such as the lifting transform, the second step in the coder construction is the design of efficient entropy coding methods for the resulting coefficients. Efficient arithmetic coding [17] of the wavelet representation of images has attracted significant interest. Recently, new coders have been proposed for lossy [18], [19] and lossless [20] compression which employ context modeling for the entropy coding of wavelet coefficients.

In this paper we first calculate the optimal predictors, in the sense of minimizing the prediction error variance of a wide-sense stationary signal, of a lifting scheme in the general n-dimensional case. Then, we apply these optimal predictor filters with corresponding update filters for the lossless compression of still images using first quincunx sampling and then simple row–column sampling. In each case, and in order to improve the results wherever the wide-sense stationarity assumption is invalid, the efficiency of the linear predictors is enhanced by nonlinear means, namely by directional postprocessing in the quincunx case, and by adaptive-length postprocessing in the row–column case. Both methods are seen to perform well. In the latter case, in particular, the resulting filter bank in effect adapts to the features of the area of the image being processed and in this way it is shown to achieve superior performance than even the most efficient known lossless compression methods capable of progressive transmission. Furthermore, we develop efficient modeling contexts which in combination with a local statistical analysis and error feedback result in superior arithmetic coding of transform coefficients.

This paper is organized as follows. Section II derives the optimal prediction filters for use with lifting for the general multidimensional case. In Section III, filters for a quincunx based

Manuscript received December 15, 1998; revised June 13, 2000. This work was supported by the Greek GSRT project PANORAMA and the European CEC IST Project INTERFACE. Part of this work was presented in [1]. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

N. V. Boulgouris and M. G. Strintzis are with the Information Processing Laboratory, Department of Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece and also with the Centre for Research and Technology—Hellas/Informatics and Telematics Institute, Thessaloniki 54639, Greece (e-mail: strintzi@eng.auth.gr).

D. Tzovaras is with the Institute of Informatics and Telematics, Thessaloniki 54006, Greece.

Publisher Item Identifier S 1057-7149(01)01180-0.

1057–7149/01$10.00 © 2000 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

decomposition are derived whose performance is enhanced by a nonlinear, direction-sensitive implementation described in Section IV. In Section V the optimal prediction filters for separable lattices are determined and synthesis filters are derived for the optimal production of intermediate images in progressive coding. A class of separable adaptive lifting transforms which enhances the performance of the filters in Section V, is presented in Section VI. An entropy coder for the coding of the resulting wavelet coefficients of Sections IV and VI, is described in Section VII. Section VIII contains experimental results and finally conclusions are drawn in Section IX.

II. Lifting Using Optimal Prediction Filters

A wavelet transform in its simplest form is obtained by filtering an image with a filter bank such as that described in Fig. 2 composed of analysis filters h_k followed by subsampling by sampling matrix $\mathbf{M}$, upsampling, and synthesis filters g_k. Exact recovery of the initial image is possible if proper relationships hold between the filters h and g [21].

In wavelet-based image representation the initial image is decomposed into a coarse image and a high pass image containing the detail information. By iterating the procedure described above to the lowpass channel, a logarithmically split filter-bank is formed consisting of wavelet coefficients at different scales and a lowpass image of very small dimensions.

The decorrelation efficiency of such a filter bank is greatly affected by the multiplicities N_k of the zeroes at π of $h_k(e^{jw})$. Thus, some of the main goals in the design of wavelet coders are maintenance of the perfect reconstruction property and proper selection of N_k. A very efficient way to accomplish this is the implementation of the filter bank using the "lifting scheme." In its most general form [10] this consists of splitting the image in separate components, estimating components from others and subsequently adding to components filtered versions of other components. The step associated with estimating the intensity of a coefficient is usually termed *prediction* whereas the step associated with smoothing the coefficients on which the initial prediction is based using the prediction errors is called *update*. If applied for lossless coding, the output of the predictor and update filters must be rounded before the addition to the corresponding components. For a first analysis, however, we shall ignore the effects of rounding.

A specific form of a filter bank coder based on a lifting scheme was proposed in [13] and is depicted in Fig. 1. In this, if $\mathbf{M}$ is the sampling matrix, $M = \det\mathbf{M}$ is the number of its polyphase components, and $\mathbf{r}_i$ its coset vectors, $i = 0, \cdots, M-1$. P_k denote the prediction and U_k the update filters $k = 0, \cdots, M-1$. The signals x_k on each branch of the filter bank are the polyphase components of the input x

$$x_k[\mathbf{Mm}] = x[\mathbf{Mm} + \mathbf{r}_k], \quad \mathbf{m} = [m_1, \cdots, m_n]^T, \quad k = 0, \cdots, M-1. \tag{1}$$

We shall assume that the input to the filter bank is a wide-sense stationary process with autocorrelation function $R[\mathbf{s}]$ and power spectral density $\Phi(e^{j\mathbf{w}})$

$$R[\mathbf{s}] = E\{x[\mathbf{m}+\mathbf{s}]x[\mathbf{m}]\}, \quad \Phi(e^{j\mathbf{w}}) = \sum_{\mathbf{s}} R[\mathbf{s}]e^{-j\mathbf{s}^T\mathbf{w}} \tag{2}$$

where the summation sign indicates a multiple sum and $\mathbf{s} = [s_1, \cdots, s_n]^T$, $\mathbf{w} = [w_1, \cdots, w_n]^T$.

The output of the prediction filters will be M distinct wide-sense stationary subsequences [22]

$$\hat{x}_i[\mathbf{Mm}] = \hat{x}[\mathbf{Mm} + \mathbf{r}_i] = \sum_{\mathbf{k}} p_i[\mathbf{Mk}]x[\mathbf{M}(\mathbf{m}-\mathbf{k})] \tag{3}$$

where $p_i[\cdot]$ is the impulse response of the ith predictor.

The coefficients $p_i[\cdot]$ will minimize the corresponding mean-square error variance

$$e^{(i)} = E\{e^2[\mathbf{Mm}+\mathbf{r}_i]\} = E\{(\hat{x}[\mathbf{Mm}+\mathbf{r}_i] - x[\mathbf{Mm}+\mathbf{r}_i])^2\} \tag{4}$$

if the following relation holds ("orthogonality principle" [23]):

$$E\left\{\left(x[\mathbf{Mm}+\mathbf{r}_i] - \sum_{\mathbf{k}} p_i[\mathbf{Mk}]x[\mathbf{M}(\mathbf{m}-\mathbf{k})]\right) \cdot x[\mathbf{M}(\mathbf{m}-\mathbf{t})]\right\} = 0. \tag{5}$$

Thus

$$R[\mathbf{Mt}+\mathbf{r}_i] = \sum_{\mathbf{k}} p_i[\mathbf{Mk}]R[\mathbf{M}(\mathbf{t}-\mathbf{k})]. \tag{6}$$

Multiplying both sides by $e^{j\mathbf{t}^T\mathbf{M}^T\mathbf{w}}$ and summing with respect to $\mathbf{t}$ yields

$$\sum_{\mathbf{t}} R[\mathbf{Mt}+\mathbf{r}_i]e^{j\mathbf{t}^T\mathbf{M}^T\mathbf{w}} = P_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right)\overline{\Phi}\left(e^{j\mathbf{M}^T\mathbf{w}}\right) \tag{7}$$

where

$$\overline{\Phi}\left(e^{j\mathbf{M}^T\mathbf{w}}\right) = \sum_{\mathbf{m}} R(\mathbf{Mm})e^{j\mathbf{m}^T\mathbf{M}^T\mathbf{w}}. \tag{8}$$

To express $\overline{\Phi}$ in terms of Φ note the identity [24], [25]

$$c[\mathbf{k}] = \frac{1}{M}\sum_{i=0}^{M-1} e^{-j2\pi\mathbf{k}^T\mathbf{M}^{-T}\mathbf{q}_i} = \begin{cases} 1, & \text{if } \mathbf{k} = \mathbf{M}\mathbf{k}_1 \ \ \mathbf{k}_1 : \text{integer} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $\mathbf{q}_i$ are the closet vectors of $\mathbf{M}^T$ and $\mathbf{M}^{-T}$ is a short notation for $(\mathbf{M}^T)^{-1}$. Thus

$$\begin{aligned}\overline{\Phi}\left(e^{j\mathbf{M}^T\mathbf{w}}\right) &= \sum_k c[\mathbf{k}]R[\mathbf{k}]e^{j\mathbf{k}^T\mathbf{w}} \\ &= \frac{1}{M}\sum_{i=0}^{M-1} R[\mathbf{k}]e^{j\mathbf{k}^T(\mathbf{w}-2\pi\mathbf{M}^{-T}\mathbf{q}_i)} \\ &= \frac{1}{M}\sum_{i=0}^{M-1} \Phi\left(e^{j(\mathbf{w}-2\pi\mathbf{M}^{-T}\mathbf{q}_i)}\right).\end{aligned} \tag{10}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

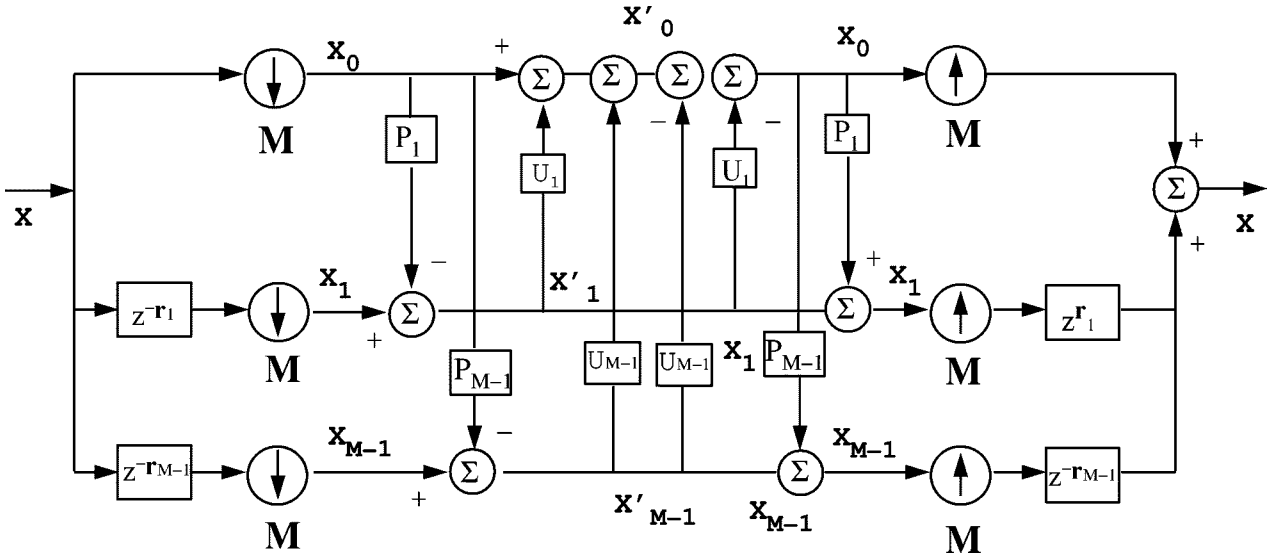


Fig. 1. M-channel lifting scheme (for simplicity, the rounding operators are not shown).

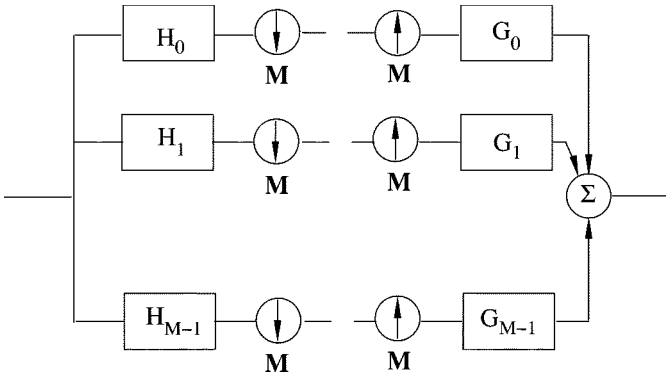


Fig. 2. Equivalent filter bank.

Finally, on multiplying both sides of (7) by $e^{j\mathbf{r}_i^T\mathbf{w}}$ and summing over i, we obtain

$$\Phi(e^{j\mathbf{w}}) = P(e^{j\mathbf{w}})\frac{1}{M}\sum_{i=0}^{M-1}\Phi\left(e^{j(\mathbf{w}-2\pi\mathbf{M}^{-T}\mathbf{q}_i)}\right)$$

where

$$P(e^{j\mathbf{w}}) = \sum_{i=0}^{M-1} e^{j\mathbf{r}_i^T\mathbf{w}} P_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) \tag{11}$$

from which the optimal predictors are immediately obtained

$$P(e^{j\mathbf{w}}) = \sum_{i=0}^{M-1} e^{j\mathbf{r}_i^T\mathbf{w}} P_i(e^{j\mathbf{M}^T\mathbf{w}}) = \frac{\Phi(e^{j\mathbf{w}})}{\overline{\Phi}(e^{j\mathbf{w}})} = \frac{\Phi(e^{j\mathbf{w}})}{\frac{1}{M}\sum_{i=0}^{M-1}\Phi\left(e^{j(\mathbf{w}-2\pi\mathbf{M}^{-T}\mathbf{q}_i)}\right)}. \tag{12}$$

It is seen therefore, that a single transfer function given by (12) may be used to describe the predictors minimizing each and every error variance $e^{(i)}$ in (4). Since the denominator of (12) is trivially seen to equal the 0th polyphase component of $\Phi(e^{j\mathbf{w}})$, the above is always, as expected, an interpolating filter, i.e., it is of the form

$$P(e^{j\mathbf{w}}) = 1 + \sum_{i=1}^{M-1} e^{j\mathbf{r}_i^T\mathbf{w}} P_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right).$$

We summarize the above results in a concluding theorem.

Theorem 1: An arbitrary lifting scheme shown in Fig. 1 is optimized, in the sense of minimizing the prediction error variances (4) for each i, $i = 1, \cdots, M-1$, if the corresponding predictor transfer functions $P_i(e^{j\mathbf{M}^T\mathbf{w}})$ are the polyphase components of the function $P(e^{j\mathbf{w}})$ found by (12).

We will now derive the expression linking the lowpass filter of a conventional filter bank with the predict and the update filters of a lifting-based implementation. The filter bank in Fig. 2 is equivalent to the filter bank of Fig. 1. Let H_k and G_k denote the transfer functions of the lowpass and highpass filters of the equivalent filter bank depicted in Fig. 2. Consider a signal $x[\mathbf{m}]$ as input to the filter bank in Fig. 2. Its Fourier transform in polyphase expansion is given by

$$X(e^{j\mathbf{w}}) = \sum_{i=0}^{M-1} e^{-j\mathbf{w}^T\mathbf{r}_i} X^i\left(e^{j\mathbf{M}^T\mathbf{w}}\right). \tag{13}$$

Let the transfer function of the kth filter be

$$H_k(e^{j\mathbf{w}}) = \sum_{i=0}^{M-1} e^{j\mathbf{w}^T\mathbf{r}_i} H_k^i\left(e^{j\mathbf{M}^T\mathbf{w}}\right), \quad k = 0, \cdots, M-1$$

where $X^i(e^{j\mathbf{M}^T\mathbf{w}})$ and $H_k^i(e^{j\mathbf{M}^T\mathbf{w}})$ denote the polyphase components of the signal and the kth analysis filter, respectively, defined as above. Then the signal in the lowpass channel of the filter bank in Fig. 2 after filtering and downsampling is

$$X'_0\left(e^{j\mathbf{M}^T\mathbf{w}}\right) = \sum_{i=0}^{M-1} H_0^i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) X^i\left(e^{j\mathbf{M}^T\mathbf{w}}\right)$$

where $H^i(\cdot)$ and $X^i(\cdot)$ are respectively the ith polyphase components of $H(\cdot)$ and $X(\cdot)$. However from (1), the ith polyphase

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

component of X is equal to the signal in the ith branch of the filter bank in Fig. 1, i.e., $X^i(e^{j\mathbf{M}^T\mathbf{w}}) = X_i(e^{j\mathbf{M}^T\mathbf{w}})$ and hence

$$X'_0\left(e^{j\mathbf{M}^T\mathbf{w}}\right) = \sum_{i=0}^{M-1} H_0^i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) X_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right). \quad (14)$$

Trivially, however from Fig. 1, we also have

$$X'_0\left(e^{j\mathbf{M}^T\mathbf{w}}\right) = X_0\left(e^{j\mathbf{M}^T\mathbf{w}}\right) + \sum_{i=1}^{M-1} X'_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) U_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right)$$

where

$$X'_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) = X_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) + P_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) X_0\left(e^{j\mathbf{M}^T\mathbf{w}}\right)$$

therefore

$$\begin{aligned} &X'_0\left(e^{j\mathbf{M}^T\mathbf{w}}\right) \\ &\quad = X_0\left(e^{j\mathbf{M}^T\mathbf{w}}\right)\left[1 - \sum_{k=1}^{M-1} P_k\left(e^{j\mathbf{M}^T\mathbf{w}}\right) U_k\left(e^{j\mathbf{M}^T\mathbf{w}}\right)\right] \\ &\qquad + \sum_{i=1}^{M-1} X_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) U_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right). \end{aligned} \quad (15)$$

From the equivalence of (14) and (15) we conclude that

$$\begin{aligned} H_0^0\left(e^{j\mathbf{M}^T\mathbf{w}}\right) &= 1 - \sum_{k=1}^{M-1} P_k\left(e^{j\mathbf{M}^T\mathbf{w}}\right) U_k\left(e^{j\mathbf{M}^T\mathbf{w}}\right) \\ H_0^i\left(e^{j\mathbf{M}^T\mathbf{w}}\right) &= U_i\left(e^{j\mathbf{M}^T\mathbf{w}}\right). \end{aligned}$$

Thus, from (13)

$$\begin{aligned} H_0(e^{j\mathbf{w}}) = 1 &- \sum_{k=1}^{M-1} P_k\left(e^{j\mathbf{M}^T\mathbf{w}}\right) U_k\left(e^{j\mathbf{M}^T\mathbf{w}}\right) \\ &+ \sum_{i=1}^{M-1} e^{j\mathbf{r}_i^T\mathbf{w}} U_k\left(e^{j\mathbf{M}^T\mathbf{w}}\right). \end{aligned} \quad (16)$$

This expression will be used in the sequel to determine the update filter so that the resulting lowpass image is smooth and thus suitable for parsimonious coding in the next step of the algorithm.

The specific case of a quincunx sampling matrix will be considered next.

III. Optimal Predictors with Quincunx Sampling

If the quincunx (see Fig. 5) sampling matrix is used

$$\mathbf{M} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (17)$$

and the separable spectral density model [26], [27] is adopted

$$R[s_1, s_2] = A\lambda_1^{|s_1|}\lambda_2^{|s_2|} \quad (18)$$

we obtain

$$\begin{aligned} &\Phi(e^{jw_1}, e^{jw_2}) \\ &\quad = \frac{A(1-\lambda_1^2)(1-\lambda_2^2)}{(1-\lambda_1 z_1)(1-\lambda_1 z_1^{-1})(1-\lambda_2 z_2)(1-\lambda_2 z_2^{-1})} \end{aligned} \quad (19)$$

where $z_k = e^{jw_k}$, $k = 1, 2$. Application of (12) yields

$$P(z_1, z_2) = 1 + \frac{\alpha_0(z_1 + z_1^{-1}) + \beta_0(z_2 + z_2^{-1})}{1 + \gamma_0(z_1 z_2 + z_1^{-1} z_2 + z_1 z_2^{-1} + z_1^{-1} z_2^{-1})} \quad (20)$$

where

$$\begin{aligned} \alpha_0 &= \frac{\lambda_1}{(1+\lambda_1^2)(1+\lambda_2^2)}, \qquad \beta_0 = \frac{\lambda_2}{(1+\lambda_1^2)(1+\lambda_2^2)}, \\ \gamma_0 &= \frac{\lambda_1\lambda_2}{(1+\lambda_1^2)(1+\lambda_2^2)}. \end{aligned} \quad (21)$$

This filtering operation may be implemented by spectral factorization, following the methods in [28] and the IIR implementation methods in [29]. Alternately, a FIR approximation to (20) may be obtained with the expansion

$$\begin{aligned} &\frac{1}{1 + \gamma_0(z_1 + z_1^{-1})(z_2 + z_2^{-1})} \\ &\quad = \sum_{k=0}^{\infty} (-\gamma_0)^k \sum_{i=0}^{k} \binom{k}{i} z_1^{k-2i} \sum_{l=0}^{k} \binom{k}{l} z_2^{k-2l} \\ &\quad = B_{0,0} + B_{1,1}(z_1 z_2 + z_1^{-1} z_2 + z_1 z_2^{-1} + z_1^{-1} z_2^{-1}) \\ &\qquad + B_{2,0}(z_1^2 + z_1^{-2} + z_2^2 + z_2^{-2}) \\ &\qquad + B_{2,2}(z_1^2 z_2^2 + z_1^{-2} z_2^2 + z_1^2 z_2^{-2} + z_1^{-2} z_2^{-2}) \\ &\qquad + B_{1,3}(z_1 z_2^3 + z_1^{-1} z_2^3 + z_1 z_2^{-3} \\ &\qquad\quad + z_1^{-1} z_2^{-3} + z_1^3 z_2 + z_1^{-3} z_2 + z_1^3 z_2^{-1} + z_1^{-3} z_2^{-1}) \\ &\qquad + B_{4,0}(z_1^4 + z_1^{-4} + z_2^4 + z_2^{-4}) + \cdots \end{aligned}$$

where

$$\begin{aligned} B_{2k,0} &= \sum_{l=k}^{\infty} \gamma_0^{2l} \binom{2l}{l} \binom{2l}{l-k}, \qquad k = 0, 1, 2, \cdots \\ B_{2,2} &= \sum_{l=1}^{\infty} \gamma_0^{2l} \binom{2l}{l} \binom{2l}{l-1}, \\ B_{1,3} &= -\sum_{l=1}^{\infty} \gamma_0^{2l+1} \binom{2l+1}{l} \binom{2l+1}{l-1}. \end{aligned}$$

Easily, (20) is seen to be approximated by

$$\begin{aligned} P(z_1, z_2) = 1 &+ (a_0 B_{0,0} + 2a_0 B_{1,1} + a_0 B_{2,0})(z_1 + z_1^{-1}) \\ &+ (\beta_0 B_{0,0} + 2\beta_0 B_{1,1} + \beta_0 B_{2,0})(z_2 + z_2^{-1}) \\ &+ (a_0 B_{1,1} + a_0 B_{1,3} + \beta_0 B_{2,0} + \beta_0 B_{2,2}) \\ &\cdot (z_1^2 z_2 + z_1^{-2} z_2 + z_1^2 z_2^{-1} + z_1^{-2} z_2^{-1}) \\ &+ (a_0 B_{2,0} + a_0 B_{2,2} + \beta_0 B_{1,1} + \beta_0 B_{1,3}) \\ &\cdot (z_1 z_2^2 + z_1 z_2^{-2} + z_1^{-1} z_2^2 + z_1^{-1} z_2^{-2}) \\ &+ a_0(B_{2,0} + 2B_{1,3} + B_{4,0})(z_1^3 + z_1^{-3}) \\ &+ \beta_0(B_{2,0} + 2B_{1,3} + B_{4,0})(z_2^3 + z_2^{-3}). \end{aligned} \quad (22)$$

For example, if only the above terms are retained, if $\lambda_1 = \lambda_2 = 0.95$ and if the above coefficients are scaled so as to sum to 1, the optimal predictor filter on quincunx sampling will be equal to

$$\begin{aligned} P(z_1, z_2) = 1 &+ 0.312(z_1 + z_1^{-1} + z_2 + z_2^{-1}) \\ &- 0.059(z_1^2 z_2 + z_1^{-2} z_2 + z_1^2 z_2^{-1} + z_1^{-2} z_2^{-1} \\ &\quad + z_1 z_2^2 + z_1 z_2^{-2} + z_1^{-1} z_2^2 + z_1^{-1} z_2^{-2}) \\ &+ 0.058(z_1^3 + z_1^{-3} + z_2^3 + z_2^{-3}). \end{aligned} \quad (23)$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

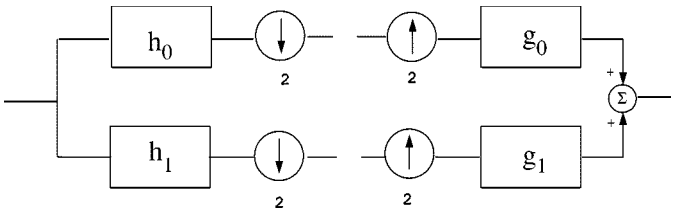


Fig. 3. Conventional implementation of a two-channel filter bank.

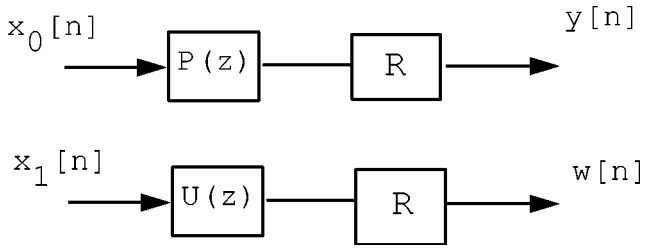


Fig. 4. Basic predictor and update units. $P(z)$ and $U(z)$ are transfer functions of linear predict and update filters and R indicates the operation of rounding to the nearest integer.

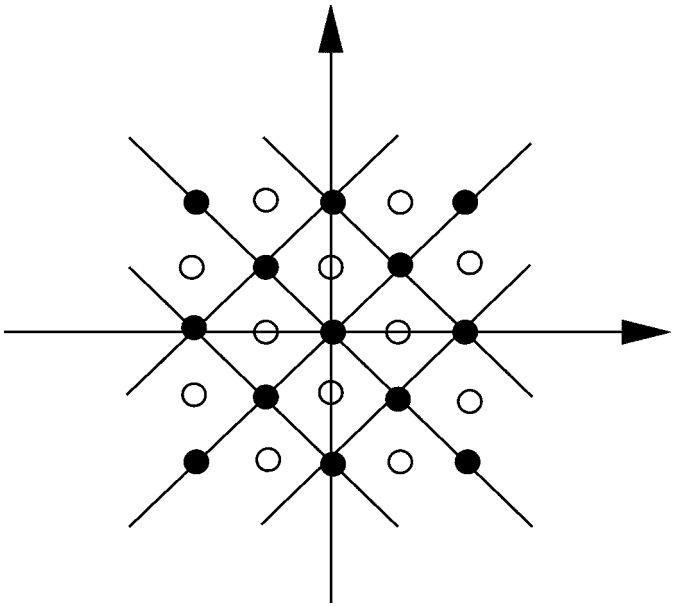


Fig. 5. Quincunx grid.

If the isotropic spectral density model

$$R[k_1, k_2] = A\lambda^{\sqrt{k_1^2+k_2^2}} \tag{24}$$

is used instead of the separable (18) model, direct approximation of (12) is difficult, since no closed form analytic expression is known for the z-transforms of (24). However, the optimal FIR filters may be determined directly from (6) by formulating a large number of equations which are then solved using least squares techniques. If specifically only the coefficients $p_i[r, s]$ with $||r|| \leq 2$, $||s|| \leq 2$, are retained, numerical solution of (6) with $\lambda = 0.95$ yields the following filter for the determination of (12)

$$\begin{aligned} P(z_1, z_2) &= 1 + e^{j\mathbf{w}^T\mathbf{r}_1} P_1\left(e^{j\mathbf{M}^T\mathbf{w}}\right) \\ &= 1 + z_2 P_1(z_1^{-1}z_2, z_1 z_2) \\ &= 1 + 0.299\,201(z_1 + z_1^{-1} + z_2 + z_2^{-1}) \\ &\quad - 0.024\,735(z_1 z_2^2 + z_1 z_2^2 + z_1^{-1}z_2^{-2} + z_1 z_2^{-2}) \\ &\quad - 0.024\,735(z_2^{-1}z_1^2 + z_2 z_1^2 + z_2^{-1}z_1^{-2} + z_2 z_1^{-2}) \\ &\quad + 0.000\,269(z_1^3 + z_1^{-3} + z_2^3 + z_2^{-3}). \end{aligned} \tag{25}$$

Both (23) and (25) are tested for each pyramid stage and the predictor producing the least error variance is used. The cost in bitrate overhead of communicating the predictor type to the decoder is negligible. The optimal prediction steps are followed by update steps using

$$U_1(z_1, z_2) = D(1 + z_1 + z_2 + z_1 z_2) \tag{26}$$

where D is a free parameter. In fact for the case of quincunx sampling $\mathbf{M}$ is given by (17), $e^{j\mathbf{M}^T\mathbf{w}} = (e^{j(\omega_1+\omega_2)}, e^{j(\omega_1-\omega_2)})$ and $\mathbf{r}_1 = [1\,0]^T$ in (16), the equivalent filter obtained is

$$\begin{aligned} H_0(e^{j\mathbf{w}}) &= 1 - P_1\left(e^{-j(w_1+w_2)}, e^{j(w_1-w_2)}\right) \\ &\quad \cdot U_1\left(e^{-j(w_1+w_2)}, e^{j(w_1-w_2)}\right) \\ &\quad + e^{jw_1} \cdot U_1\left(e^{-j(w_1-w_2)}, e^{j(w_1+w_2)}\right). \end{aligned} \tag{27}$$

The parameter D may be derived by forcing the sum of the coefficients of the 0th and the 1st polyphase component of the equivalent filter $H(\mathbf{z})$ to be equal. This constraint is equivalent to requiring that $H(e^{j\mathbf{w}})$ have a zero at $w_1 = \pi$,$w_2 = \pi$. Easily, (27) yields

$$U_1(1, 1)(P_1(1, 1) + 1) = 1. \tag{28}$$

Since a "balanced" transfer function is required, $H_0(1, 1) = 1$, and hence (27) implies $P_1(1, 1) = 1$. Thus, from (28)

$$U_1(1, 1) = \tfrac{1}{2}.$$

From which and (26), D is directly determined

$$D = \tfrac{1}{8}.$$

This result holds independently of the actual prediction filter $P(z_1, z_2)$ used as long as $P_1(1, 1) = 1$. This provides the appropriate update transfer function

$$U_1(z_1, z_2) = \tfrac{1}{8}(1 + z_1 + z_2 + z_1 z_2) \tag{29}$$

for the quincunx based decomposition. A total number of at least four decomposition levels is needed for efficient decorrelation of most grayscale images.

A nonlinear scheme, named MINT, for the enhancement of the performance of this linear optimal prediction filter, originally proposed and evaluated in [22], is described in the next section. In the enhanced coder, named MINT-U, prediction steps using the statistically optimal prediction filters derived above are followed by nonlinear processing and appropriate update steps.

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

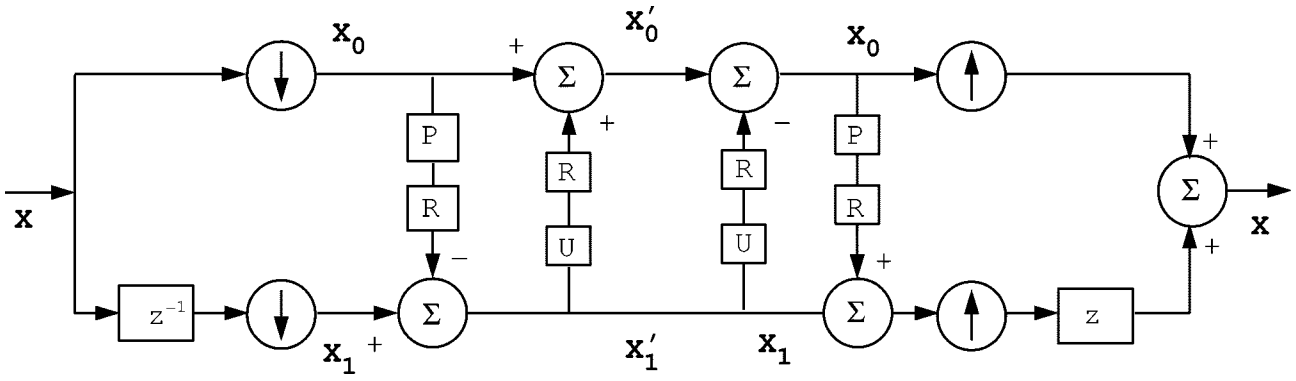


Fig. 6. Simple lifting scheme using the predictor and update units of Fig. 4.

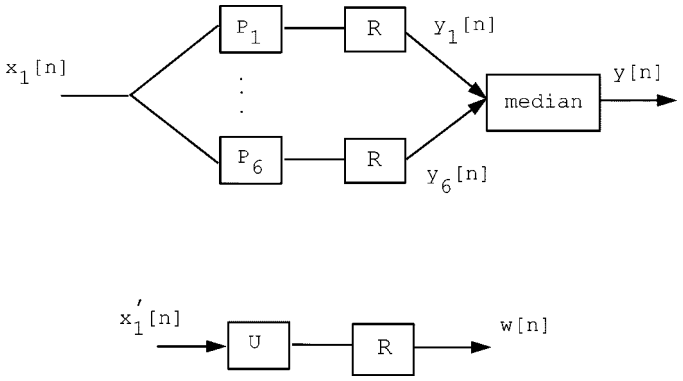


Fig. 7. Proposed design of predictor and update units.

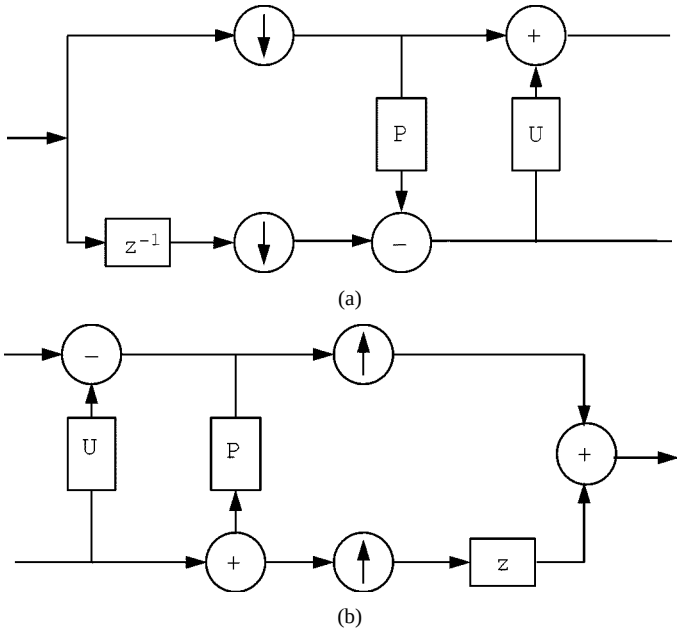


Fig. 8. Implementation of a two-channel filter bank using nonlinear prediction steps. (a) Analyzing bank and (b) synthesizing bank.

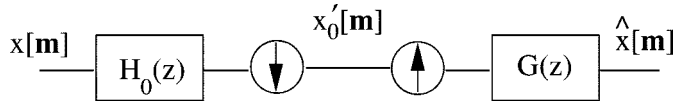


Fig. 9. General system for the optimization of intermediate image quality.

IV. Nonlinear Directional Prediction

The preceding section was devoted to the determination of the optimal linear prediction filters. In the present section we shall develop nonlinear adaptations of the optimal linear filters which are orientation-sensitive, so as to adapt to the local image features and thus further improve the prediction. More specifically we shall employ the *median hybrid* filters [30]–[32] as combinations of median and linear filters

$$y[i] = med\left(y^{(k)}[i],\, i = 1, \cdots, M\right)$$

where $y^{(k)}[i],\, i = 1, \cdots, M$ are the outputs of the linear FIR or IIR filters $P_k(z)$.

More specifically, we propose the splitting of the optimal interpolation filter into four directional prediction filters and the selection of their median as the estimate for the current pixel. This technique originally described in [22] will be referred to as median-based minimum variance interpolation (MINT). For each pixel the four directional predictors are given by

$$P_N(z_1, z_2) = \sum_{i=-N_1}^{0} \sum_{k=-N_1}^{0} p_1(i, k) z_1^{-i} z_2^{-k}$$

$$P_E(z_1, z_2) = \sum_{i=0}^{N_1} \sum_{k=-N_1}^{0} p_1(i, k) z_1^{-i} z_2^{-k}$$

$$P_W(z_1, z_2) = \sum_{i=-N_1}^{0} \sum_{k=0}^{N_1} p_1(i, k) z_1^{-i} z_2^{-k}$$

$$P_S(z_1, z_2) = \sum_{i=0}^{N_1} \sum_{k=0}^{N_1} p_1(i, k) z_1^{-i} z_2^{-k}.$$

Following rounding, the median of the four predictors is applied as follows: the resulting four predictions are ordered according to their magnitude and then the average of the second and third is taken as the prediction. This modification in the prediction step gives the coder the flexibility to select the best predictor depending on the features present in each local area. Thus the estimate $\hat{y}(i, j)$ of $y(i, j)$ will be given by

$$\hat{y}(l, m) = med(y_N(l, m), y_E(l, m), y_W(l, m), y_S(l, m))$$

where $y_d(l, m)$ is the output of the directional filter with $d = N, E, W$ and S. These filters are often referred to as $1LM+$ filters in the [30]–[32].

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

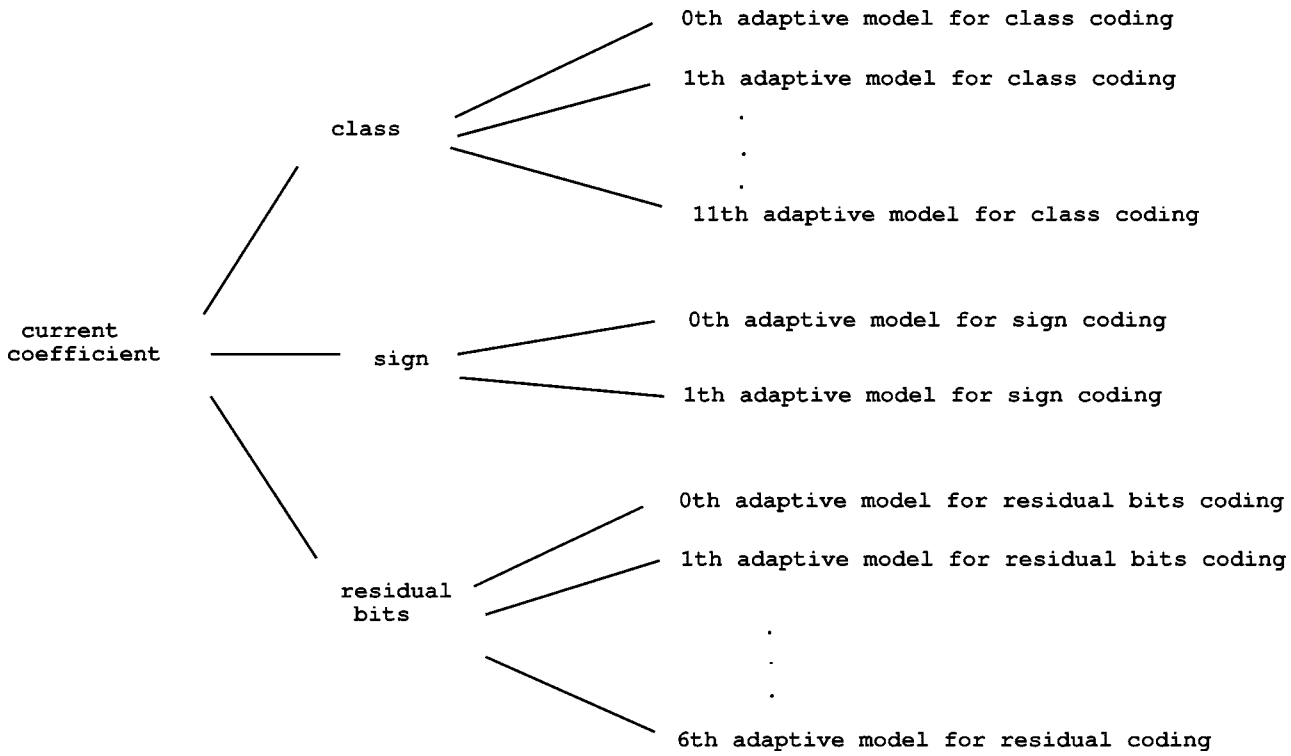


Fig. 10. Schematic description of the proposed entropy coder. The current coefficient is decomposed in three components, namely class, sign and residual bits. Twelve probability models are used for coding the class. Two probability models are used for coding the sign. Seven probability models are used for coding the residual bits.

V. Optimal Row–Column Sampling and Optimal Production of Intermediate Images in Progressive Coding

With row–column sampling where the matrices

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

are used alternatively, the problem of determining the optimal prediction reduces to the one-dimensional case, i.e., to having $M = 2$. Then, (10) and (12) degenerate to

$$\begin{aligned} \Phi(e^{jw}) &= G(e^{jw})\tfrac{1}{2}\left(\Phi(e^{jw}) + \Phi\left(e^{j(w-\pi)}\right)\right) \\ P(e^{jw}) &= 1 + e^{-jw}P_1(e^{j2w}) \\ &= \frac{\Phi(e^{jw})}{\overline{\Phi}(e^{jw})} = \frac{\Phi(e^{jw})}{\frac{1}{2}(\Phi(e^{jw}) + \Phi(-e^{jw}))}. \end{aligned} \tag{30}$$

For this, it is necessary to specify the spectral densities describing the class of images which are of interest. As before, the optimal prediction filter will be determined using the separable spectral density model [26], [27]

$$R[s] = A\lambda^{|s|}$$

hence

$$\Phi(e^{jw}) = \frac{A(1-\lambda^2)}{(1-\lambda z)(1-\lambda z^{-1})} \tag{31}$$

where $z = e^{jw}$; application of (12) yields immediately

$$1 + zP_1(z^2) = 1 + \frac{\lambda}{1+\lambda^2}(z + z^{-1}) \tag{32}$$

whence, for the first level of the filter bank

$$P_1(z^2) = \frac{\lambda}{1+\lambda^2}(1 + z^{-2}).$$

From (16), we have

$$H_0(z) = 1 - P_1(z^2)U_1(z^2) + zU_1(z^2).$$

A choice for the update filter which has produced satisfactory results [12] is $U_1(z) = D(1+z)$ where D is a constant. This choice will not influence the investigation of the optimal prediction filter. As discussed earlier, it is desirable that $H_0(z)$ have at least one zero at $\omega = \pi$, i.e., at $z = -1$

$$\begin{aligned} H_0(z) = 0|_{z=-1} &\Rightarrow H_0^0(z^2) + z^{-1}H_0^1(z^2) = 0|_{z=-1} \\ &\Rightarrow H_0^0(1) - H_0^1(1) = 0 \Rightarrow H_0^0(1) = H_0^1(1). \end{aligned}$$

Therefore, if we also require that the analysis filter is balanced, i.e., $H_0(1) = 1$

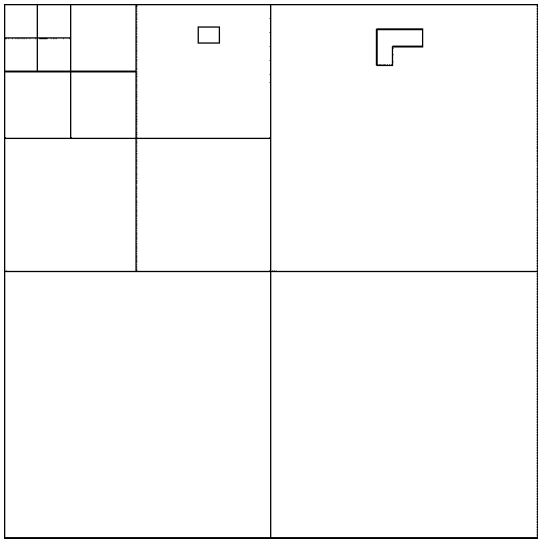
$$\Rightarrow H_0^1(1) = 0.5 \Rightarrow D(1+1) = 0.5 \Rightarrow D = \tfrac{1}{4}.$$

Thus, independently of the prediction filter $P(z)$ used, a reasonable choice for a update filter of length two is

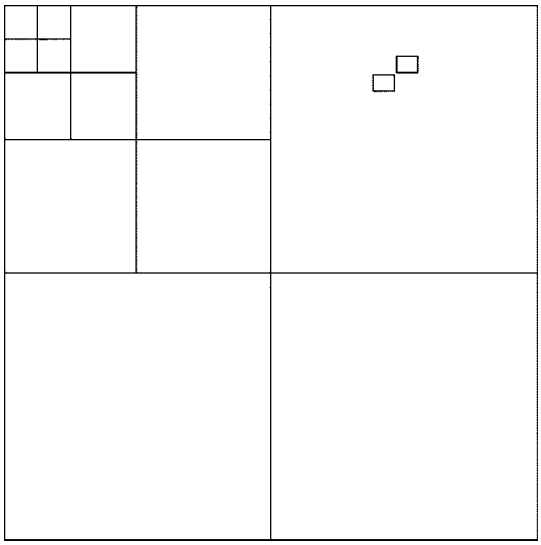
$$U_1(z) = \tfrac{1}{4}(1+z). \tag{33}$$

The above prediction and update filter combination was identified in [12] as the (2, 2) integer wavelet having two zeros at π of the analyzing and synthesizing lowpass filter transfer functions (see Fig. 8). Other filter banks of the same class are also studied in [12] and are seen to have comparable and often better overall performance. This performance depends on the type of image being coded. In the ensuing section an adaptive technique for the implementation of reversible integer transforms will be

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.



(a)

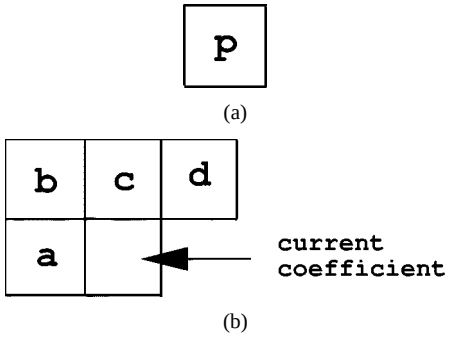


(b)

Fig. 11. Contexts used for (a) class (b) sign.

presented using several alternative filters as well as the above optimal predictor.

It is important to note, that this filter bank was chosen solely so as to minimize the prediction error variance and hence, that the images produced by the corresponding lowpass band of the filter bank are not necessarily close approximations of the original image $x[m_1, m_2]$ as one would very much desire in progressive transmission. To optimize the quality of the intermediate image it suffices to pass at each level the lowpass image $x'_0[\mathbf{m}]$ through an upsampler and a synthesis filter given by the equation shown at the bottom of the page [33] (see Fig. 9) where



(a)

(b)

Fig. 12. (a) Parent coefficient and (b) adjacent pixels used for class coding.

$H_0(e^{j\mathbf{w}})$ is the analysis filter (16) in the equivalent formulation of Fig. 3. In the case of a quincunx matrix $\mathbf{M}$ this reduces to the equation shown at the bottom of the next page and in the simple one-dimensional (1-D) case

$$G_0(z) = \frac{2H_0(z^{-1})\Phi(z)}{H_0(z)\Phi(z)H_0(z^{-1}) + H_0(-z)\Phi(-z)H_0(-z^{-1})} \tag{34}$$

where $H_0(z)$ is the analysis filter of the equivalent formulation in Fig. 3

$$H_0(z) = 1 - P_1(z^2)U_1(z^2) + zU_1(z^2).$$

In fact, it has been shown in [33] that the above choice of the synthesis filter minimizes the error variance $E\{(x[m]-\hat{x}[m])^2\}$ and hence makes the intermediate output $\hat{x}[m]$ as similar to the original as is possible by means of linear postfiltering.

VI. A Class of Adaptive Nonlinear Lifting Transforms

Precisely as linear prediction based on quincunx sampling may be improved by the direction-sensitive nonlinear postfiltering in Section III, simple row–column sampling may be very considerably improved if adaptive, nonlinear variations of the optimal filters are used. The approach proposed in this Section employs nonlinear prediction steps to formulate a lifting scheme which adapts to the features of the area of the image being processed. Instead of using a single filter, three filters with different lengths are used. For the update step a single filter is used. Specifically, the filter banks referred to as (6,2),(4,2) in [12] are used, as well as the nominally optimal predictor (2,2) method examined in Section III. These filter banks have correspondingly 6,4,2 vanishing moments of the analyzing high-pass filter and two vanishing moments of the analyzing low pass filter. After their factorization into lifting steps [11], all three filter banks employ the same update step. If $P_1(z)$ denote the transfer functions of the linear prediction filters and $U_1(z)$ the transfer

$$G_0(e^{j\mathbf{w}}) = \frac{MH_0(e^{-j\mathbf{w}})\Phi(e^{j\mathbf{w}})}{\sum_{k=0}^{M-1} H_0\left(e^{j(\mathbf{w}+2\pi\mathbf{M}^{-T}\mathbf{q}_k)}\right)\Phi\left(e^{j(\mathbf{w}+2\pi\mathbf{M}^{-T}\mathbf{q}_k)}\right))H_0\left(e^{-j(\mathbf{w}+2\pi\mathbf{M}^{-T}\mathbf{q}_k)}\right)}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

functions of the update filters (see Fig. 6), the filter banks are described by

(2,2)

$$P_1(z) = \tfrac{1}{2}(1 + z^{-1})$$
$$U_1(z) = \tfrac{1}{4}(1 + z) \tag{35}$$

(4,2)

$$P_1(z) = \tfrac{9}{16}(1 + z^{-1}) - \tfrac{1}{16}(z^{-2} + z)$$
$$U_1(z) = \tfrac{1}{4}(1 + z) \tag{36}$$

(6,2)

$$P_1(z) = \tfrac{75}{128}(1 + z^{-1}) - \tfrac{25}{256}(z^{-2} + z) + \tfrac{3}{256}(z^{-3} + z^2)$$
$$U_1(z) = \tfrac{1}{4}(1 + z). \tag{37}$$

In [12] it was pointed out that transforms having more analyzing vanishing moments perform better when applied to smooth images but their performance suffers for images containing many edges. Our intent here is to form a transform capable of universally exceptional performance for all types of images being transformed. In order to achieve this, each one of the three filters is first split into two parts (noncausal and causal) providing six different filters $P_i(z)$ and six corresponding predictions p_i, $i = 1, \cdots, 6$, i.e.,

$$P_1^1(z) = 1 \quad P_1^2(z) = z^{-1} \quad P_1^3(z) = \tfrac{9}{8}\, z^{-1} - \tfrac{1}{8}\, z^{-2}$$
$$P_1^4(z) = \tfrac{9}{8} - \tfrac{1}{8}\, z \quad P_1^5(z) = \tfrac{75}{64} - \tfrac{25}{128}\, z + \tfrac{3}{128}\, z^2$$
$$P_1^6(z) = \tfrac{75}{64}\, z^{-1} - \tfrac{25}{128}\, z^{-2} + \tfrac{3}{128}\, z^{-3}.$$

Following rounding, the median of the six predictors is applied as follows: the resulting six predictions are ordered according to their magnitude and then the average of the third and fourth is taken as the prediction. This modification in the prediction step gives the coder the flexibility to select the best predictor depending on the features present in each local area. The prediction step is followed by the update step which is the same as that used in the three filter banks and in Section V, the proposed transform is illustrated in Fig. 7. A total number of at least four levels of lowpass image decomposition is generally needed for satisfactory coding performance.

Despite their excellent decorelating performance, which will be demonstrated in the experimental results section, wavelet transforms are not able to remove all existing redundancy in a still image. This can be easily confirmed by inspection of the wavelet representation of an image. It is seen that wavelet coefficients residing in edge areas have high variance at all scales of the decomposition. These coefficients are the most expensive to code. In the ensuing section we shall describe a novel approach for the exploitation of this residual redundancy during the entropy coding procedure.

VII. Conditional Arithmetic Coding of Transformed Images

For the coding of the coefficients produced by the nonlinear transform of Sections IV and VI, an entropy coder combining many features was implemented. The sequence of wavelet coefficients is first partitioned into a number of subsequences which consist of symbols having similar statistics. Each subsequence is coded using a separate arithmetic coder, i.e., a variety of adaptive probability models are employed. The lowpass image is first decorrelated using linear DPCM and then coded using a separate adaptive model. Each coefficient is represented by the class in which it belongs, its sign and the residual bits specifying its exact magnitude. The classes used are shown in Table II. For example a coefficient with a value of -80 is represented by the triad $(10, -, 16)$. The first number in the triad is the class number. Since the coefficient belongs to the tenth class, it is valued between 64 and 127. The symbol "$-$" is its sign and the number 16 is the residual difference between the number being depicted and the lowest number in its class: $16 = 80 - 64$.

This representation is similar to that used in JPEG [34] and the S+P transform [35] for lossless coding, as well as in [18] for lossy compression. The whole range of the coefficients is divided into several classes. Due to the fact that the values of wavelet coefficients are concentrated around zero, more classes are allotted to near-zero magnitudes (Table II). Some classes (0,1,2,3) include only one coefficient magnitude. This means that a coefficient belonging to such a class does not require residual bits for its representation since the class itself identifies the exact magnitude of the coefficient.

Each coefficient in the highpass bands is conditioned using the values of adjacent coefficients and the parent coefficient, i.e., the coefficient lying in the lower scale and in the same spatial orientation as shown in Fig. 11. The causal contexts used presume that the exact value of these wavelet coefficients is known, i.e., all class, residual and sign bits of the previously coded coefficients are known. This means that the decoder while decoding the current coefficient has fully recovered the past coefficients. In this way, past coefficients that were used as a conditioning context during encoding, are also *explicitly* known by the decoder during the decoding process. This fact makes the formation of modeling contexts more flexible in comparison to lossy coders since in the later case only quantized and not the original coefficient values are known.

An arithmetic coder [17] achieves significant compression by transmitting the more probable symbols in fewer bits than the less probable ones. For example, the model may assign a predetermined probability to each possible symbol. These probabilities may be determined by counting frequencies in

$$G_0(z_1, z_2) = \frac{2H_0(z_1^{-1}, z_2^{-1})\Phi(z_1, z_2)}{H_0(z_1, z_2)\Phi(z_1, z_2)H_0(z_1, z_2) + H_0(-z_1, -z_2)\Phi(-z_1, -z_2)H_0(-z_1, -z_2)}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF THE PROPOSED AL, MINT-U (M-U) ADAPTIVE LIFTING TRANSFORMS WITH THE HINT, MORE, S+P AS WELL AS SEVERAL INTEGER WAVELET TRANSFORMS. WEIGHTED ENTROPIES OF TEST IMAGES ARE REPORTED

Image	Raw	HINT	MORE	S+P	2,2	4,2	2,4	4,4	2+2,2	6,2	M-U	AL
Lenna	7.44	4.53	4.30	4.33	4.35	4.30	4.36	4.30	4.30	4.31	4.33	4.29
F16	6.79	5.37	5.07	5.19	5.14	5.17	5.17	5.17	5.16	5.20	5.14	5.12
Crowd	6.78	4.59	4.22	4.29	4.35	4.26	4.39	4.28	4.27	4.26	4.28	4.24
Moon	6.71	5.12	4.98	5.00	5.00	5.01	5.02	5.00	5.00	5.02	4.99	4.99
Girl	6.42	4.80	4.62	4.65	4.65	4.62	4.67	4.62	4.62	4.63	4.65	4.61
Mall	6.99	5.49	5.17	5.22	5.27	5.21	5.31	5.22	5.20	5.22	5.23	5.20
Pentagon	6.52	5.57	5.32	5.38	5.40	5.38	5.42	5.38	5.38	5.39	5.34	5.37
Peppers	7.59	5.08	4.61	4.67	4.61	4.61	4.62	4.60	4.61	4.63	4.61	4.58
Couple	5.96	4.48	4.21	4.25	4.19	4.20	4.21	4.20	4.19	4.22	4.22	4.17
Bridge	7.66	6.00	5.88	5.88	5.85	5.85	5.87	5.85	5.84	5.87	5.89	5.84
Barbara	7.46	5.17	5.25	4.94	5.00	4.86	4.97	4.83	4.86	4.81	5.26	4.88
Zelda	7.33	4.05	3.92	3.87	3.85	3.83	3.85	3.81	3.83	3.84	3.95	3.81
Gold	7.53	4.84	4.73	4.73	4.68	4.68	4.70	4.68	4.67	4.70	4.75	4.67
GirlII	7.29	4.31	4.05	4.06	4.08	4.00	4.10	3.99	3.99	3.98	4.07	3.97
Hotel	7.55	4.97	4.69	4.97	4.71	4.70	4.73	4.70	4.70	4.72	4.72	4.67
Boats	7.09	4.51	4.31	4.26	4.23	4.19	4.24	4.19	4.18	4.20	4.36	4.17
US	4.88	3.78	3.62	3.78	3.71	3.87	3.78	3.89	3.84	3.94	3.77	3.72

representative samples of the input source to be transmitted. Such a *fixed* model is communicated in advance to both the encoder and decoder, after which it is used for many images. Alternatively, the probabilities the model assign may change as each symbol is transmitted, based on the symbol frequencies seen *so far* in this message. In such an *adaptive model* there is no need for a representative sample of input data, because each message is treated as an independent unit, starting from scratch. The encoder's model changes with each symbol transmitted, and the decoder's changes with each symbol received, in precisely the same manner. Adaptive models are used in all our experiments.

Splitting of the information of a coefficient in three components, namely class, residual bits and sign, requires the use of different adaptive arithmetic coders (adaptive models) for the arithmetic coding of each component, since it is to be expected that the symbols corresponding to the three components have different statistics. The adaptivity of the coder means that initially all symbols are considered of having equal probability. During encoding, however, the probability tables are updated and soon converge to the actual statistics of the source maximizing the coding efficiency of the arithmetic coder. In order to further enhance performance, we used many adaptive models for coding each of the three components in which a coefficient is split.

In each quadrant, the wavelet coefficients are visited in a lexicographical order. For each coefficient, a "context" C_j is derived based on the values of its adjacent coefficients which have already been visited and its parent coefficient. For example, for coding the class of a coefficient, the index of the adaptive model that will be used is derived based on its causal neighborhood and then the corresponding adaptive model is used for the coding of the class. In this way the activity in the vicinity of the current coefficient provides some information about the magnitude of the current coefficient. By taking this into consideration, coefficients of different activity are treated differently by the encoder, by being fed to different adaptive coders. This higher order approach reduces the eventual errorless encoding cost below the first order entropy of the initial coefficient sequence. Similar conclusions hold for residual bit and sign coding.

A. Coding the Class

Using the conditioning contexts described above and depicted in Fig. 11(a), the adaptive probability model which will be used for the coding of the current coefficient is determined. This is done as follows:

$$C_p = CLASS(A) \tag{38}$$

where the parameter A equals

$$A = \frac{0.5|p| + 1.5\dfrac{|a| + |b| + |c| + |d|}{4}}{2} \tag{39}$$

where p denotes the parent coefficient [Fig. 12(a)] and a,b,c,d the four causal, adjacent coefficients of the current coefficient [Fig. 12(b)]. Equation (38) gives the index of the adaptive probability model which will be used for the coding of the class of the current coefficient. The weighting factors in (39) were

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

TABLE II
CLASSES USED FOR THE CODING OF THE MAGNITUDE INFORMATION OF COEFFICIENTS

Class	magnitude
0	0
1	1
2	2
3	3
4	4-5
5	6-7
6	8-11
7	12-15
8	16-31
9	32-63
10	64-127
11	128-255

chosen heuristically, on the basis of experimentation. In [36] it was shown that features in different bands do not have the same orientation. However, in our coder we used the same context, i.e., the same causal set of coefficients, for all bands since we found that taking into account different feature orientations, by forming different conditioning contexts for different bands, does not yield significant profit.

For example consider, as before, a coefficient valued -80. This coefficient is represented by the triad $(10, -, 16)$ The first number in the triad is the class number. Thus, a coefficient belonging to the 10th class is valued between 64 and 127. The symbol "$-$" is its sign and the number 16 is the residual difference between the number being depicted and the lowest number in its class: $80 = 64 + 16$. If/then equation (38) gives $CLASS(70) = 10$, the tenth arithmetic model will be used to code the class of the current coefficient. In a similar way, all coefficient classes (0–11) are coded using the adaptive model whose index is derived using (38).

The coding of the class parameter can be further optimized by using an error feedback technique, similar to that in [20], based on local statistical analysis. Specifically, equation (38) provides an index of the adaptive model that will be used for coding the class. If the prediction errors are taken into consideration then the efficiency of class coding may be improved. In order to achieve this, we consider the output of (38) and (39) as preliminary information for the class of the current coefficient. Let e_K be the difference of the predicted value (39) and the actual coefficient magnitude at the Kth occurrence of the class. Let also $mean\_error[C_p(N)] = (1/N)\sum_1^N e_K$ be the average of prediction errors e_K for class C_p for $K = 1$ to $K = N$. This error may be computed iteratively as follows:

$$mean\_error[C_p(N+1)] = \frac{N}{N+1} mean\_error[C_p(N)] + \frac{e_{N+1}}{N+1}. \quad (40)$$

In subsequent steps the prediction A will be biased by the mean error corresponding to the class where A belongs. In this way, the parameter A may be evaluated again as

$$\hat{A} = A - mean\_error[C_p].$$

In fact, as experimentally verified, an even better estimation of A, yielding increased compression is afforded by (41), shown at the bottom of the page.

The eventual index of the adaptive model that will be used is given by

$$\hat{C}_p = CLASS(\hat{A}). \quad (42)$$

The procedure described above is summarized in the following pseudocode:

- Compute A using (39).
- Find C_p using (38).
- Compute $\hat{A}$ using (41).
- Find $\hat{C}_p$ using (42).
- Update bias using (40)

In our practical coder the parameter N was chosen to be smaller than ten. In this way more recent symbols are given a higher weight in the calculation of the mean error than symbols in the more remote past.

B. Coding the Residual Bits

The coding of the residual bits uses a different adaptive arithmetic model for each class. Since the decoder is aware of the class in which the current coefficient belongs, it decodes the residual bits using the adaptive model which has been formulated for the decoding of the residual bits of this class.

C. Coding the Sign

Efficient coding of the signs is probably the most challenging of the three tasks. We code the signs coefficients using a very simple context selector and two adaptive probability models. We used the available causally adjacent pixels lying on the horizontal and vertical directions [see Fig. 11(b)]. If the sum of these two coefficients is positive, we use the first adaptive arithmetic coder. Otherwise, we use the second arithmetic coder for the coding of the current sign. This intuitive approach produces a

$$\hat{A} = A - \frac{mean\_error[C_p - 1] + 2mean\_error[C_p] + mean\_error[C_p + 1]}{4} \quad (41)$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

TABLE III
COMPARISON OF THE PROPOSED ADAPTIVE LIFTING WITH CONDITIONAL ARITHMETIC (ALCA) CODING WITH THE JPEG-LS (J-LS), THE S+P CODER AS WELL AS INTEGER WAVELET TRANSFORM CODERS. EXACT BIT RATES ARE REPORTED. THE ENTROPY CODER DESCRIBES IN SECTION VII WAS USED WITH ALL INTEGER WAVELET TRANSFORM CODERS APART FROM THE S+P AND THE AL TRANSFORM WHICH WERE ALSO USED WITH THE S+P ENTROPY CODER

Image	J-LS	S+P	SPCA	2,2	4,2	2,4	4,4	2+2,2	6,2	M-UCA	ALSP	ALCA
Lenna	4.24	4.17	4.13	4.15	4.11	4.16	4.11	4.11	4.12	4.15	4.14	**4.10**
F16	**4.70**	4.88	4.83	4.80	4.81	4.82	4.81	4.81	4.83	4.80	4.83	4.79
Crowd	**3.91**	4.00	3.97	4.03	3.95	4.07	3.96	3.95	3.94	3.96	3.96	3.93
Moon	5.08	5.05	5.01	5.00	5.01	5.02	5.00	5.01	5.02	**4.99**	5.04	5.00
Girl	4.62	4.56	4.52	4.52	4.50	4.54	**4.49**	4.50	4.50	4.53	4.54	**4.49**
Mall	4.94	4.97	4.94	5.00	**4.92**	5.04	4.93	4.93	4.93	4.97	4.96	**4.92**
Pentagon	5.28	5.31	5.28	5.31	**5.27**	5.32	**5.27**	**5.27**	5.28	5.29	5.31	**5.27**
Peppers	4.51	4.58	4.50	4.46	4.46	4.46	4.45	4.46	4.47	4.48	4.51	**4.44**
Couple	**3.77**	3.97	3.93	3.90	3.90	3.92	3.90	3.90	3.92	3.93	3.92	3.88
Bridge	5.79	5.81	5.76	5.75	5.75	5.77	5.75	**5.74**	5.76	5.81	5.79	**5.74**
Barbara	4.86	4.55	4.48	4.55	4.45	4.53	**4.42**	4.45	**4.42**	4.77	4.54	4.46
Zelda	3.89	3.83	3.77	3.74	3.73	3.73	**3.70**	3.73	3.73	3.85	3.77	3.71
Gold	4.48	4.56	4.49	4.46	4.46	4.47	4.46	4.45	4.47	4.53	4.50	**4.44**
GirlIII	3.93	3.96	3.90	3.93	3.84	3.94	3.84	3.84	3.83	3.94	3.89	**3.83**
Hotel	**4.38**	4.53	4.48	4.44	4.44	4.45	4.44	4.43	4.46	4.49	4.46	4.41
Boats	3.93	4.03	3.96	3.94	3.92	3.95	**3.90**	3.91	3.92	4.08	3.96	**3.90**
US	**2.63**	3.16	3.11	3.07	3.18	3.13	3.19	3.14	3.22	3.08	3.13	3.07

small gain in coding. This gain is due to the fact that the partitioning of the sign stream, using the aforementioned rule, produces two streams which have different statistics. Note, that since adaptive arithmetic coders are used, these statistics need not be computed, stored or transmitted. The adaptive arithmetic coder simply adapts to the statistics of each of the two input streams.

As a whole, our coder uses 12 adaptive probability models (0–11) for the coding of the class of the current pixel, seven models for the coding of the residual bits and two models for the coding of the sign. It is important to note that all these models are adaptively updated while coding and decoding. This means that the arithmetic coders learn the statistics of the input sequences. The adaptive models are reset at different levels of the pyramid. The entropy coder is schematically described in Fig. 10.

VIII. EXPERIMENTAL RESULTS

The efficiency of the proposed methods for lossless coding was evaluated using a variety of grayscale images. The MINT-U method was implemented using the prediction filters (23) and (25), as described in Section III, and the update filter (29). The separable AL transform was implemented using prediction filters having 6, 4, and 2 vanishing moments and the update filter given by equation (33). A maximum number of six levels of decomposition was performed for all images. Note that in both algorithms, adaptation is always based on information that is available during the decoding process and the scheme is perfectly reversible without the need for other information.

The evaluation is based on the error image weighted first order entropy and the exact lossless compression rates. The algorithms compared were a collection of lossless coding algorithms having the capacity for progressive transmission including the HINT method [3], [7], [37], the recently proposed MORE coder [9], the S+P method [35] and a variety of integer wavelets [12], methods which are widely considered to represent the state-of-the-art in this area. Final weighted first order entropies are reported in Table I. As seen from this table, MINT-U in some cases and AL in almost all cases, lead to better results than all other methods. AL appears to be generally more efficient than MINT-U. This contradicts to the intuitive expectation that nonseparable prediction should be more efficient than the separable one. Indeed, the optimal prediction performed by MINT-U on a quincunx lattice is generally more efficient. However, the subsequent update step, smoothing the zeroth polyphase component of the image, is less efficient and the image is less smooth than that of the separable scheme. It appears that in most cases the better smoothing achieved by the separable scheme cancels out the better prediction of the optimal nonseparable scheme. Additionally, in the nonseparable case, the error terms, once calculated, are not further decomposed as in the separable case where row–column filtering is applied on both low-pass and error terms. We regard this as a possible method for the improvement of the performance of the nonseparable scheme and improving the results of our method.

In conjunction with the efficient conditional arithmetic coder of Section VII, the novel decorrelation algorithms MINT-U

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

and AL form complete lossless compression methods termed MINT-UCA and ALCA respectively. The lossless compression performance of these methods in terms of exact bitrates is reported in Table III. Rates obtained using the adaptive lifting (AL) transform and the S+P entropy coder are also included (ALSP). The S+P transformed images were encoded with both the S+P entropy coder and the new entropy coder described in this paper (this coder is termed SPCA in Table III). It is shown that the new entropy coder consistently outperfroms the S+P entropy coder for all images. As seen from Table III, both proposed methods outperform the state-of-the-art S+P coder. ALCA in particular is seen to outperform the lossless compression algorithms capable of progressive transmission that are included in our comparison. The proposed ALCA coder is also completive with the JPEG-LS[1] coder which is based on predictive methods and therefore does not afford preview and progressive transmission capability. Other wavelet coders, e.g., [38], produce embedded streams and achieve very efficient lossless compression using pattern recognition methods for context classification during the context arithmetic coding. However, this comes at the cost of significantly increased complexity for the entropy coder.

IX. Conclusions

In this paper, the optimal predictors of a lifting scheme in the general n-dimensional case were calculated and applied with corresponding update filters for the lossless compression of still images using first the quincunx sampling matrix and then simple row–column sampling. In each case, the efficiency of the linear predictors was enhanced by directional nonlinear postprocessing in the quincunx case, and by adaptive-length nonlinear postprocessing in the row–column case. Both methods (MINT-U and AL respectively) are seen to perform well in the sense of leading to lower first order entropies than other methods, with AL being more efficient than MINT-U in most cases.

The exploitation of redundancy in wavelet transformed images during entropy coding of the MINT-U and AL methods was also investigated. Different contexts were employed for coding the class, residual bits and sign information. Correspondingly, a number of different adaptive models were used for coding each of the three components. Experimental results were obtained by applying the proposed methods to a large number of images. These evaluated the final bitrate needed for the implementation of the new coders using the special efficient entropy coder of Section VII. The performance of both schemes was seen to be competitive with that of state-of-the-art coders and ALCA in particular was shown to outperform other known algorithms for lossless image coding with intermediate view capability in progressive transmission.

References

[1] N. V. Boulgouris and M. G. Strintzis, "Reversible multiresolution image coding based on adaptive lifting," in *Proc. IEEE Int. Conf. Image Processing*, Kobe, Japan, Oct. 1999.

[1]The JPEG-LS Reference Encoder - V.1.00 was used for obtaining the results presented in this paper. JPEG-LS executables can be downloaded from http://www.hpl.hp.com/loco/

[2] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo image sequences using joint 3-D motion/disparity compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 312–328, Apr. 1997.
[3] P. Roos and M. A. Viergener, "Reversible 3-D decorrelation of medical images," *IEEE Trans. Med. Imag.*, vol. 12, pp. 413–420, Sept. 1993.
[4] N. V. Boulgouris, A. Leontaris, and M. G. Strintzis, "Wavelet compression of 3D medical images using conditional arithmetic coding," in *Proc. IEEE Int. Symp. Circuits Systems*, Geneva, Switzerland, May 2000.
[5] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Commun.*, vol. 45, pp. 437–444, Apr. 1997.
[6] M. J. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: A low complexity, context-based lossless image compression algorithm," in *Proc. DCC Data Compression Conf.*, Snowbird, UT, 1996, pp. 140–149.
[7] P. Roos, M. A. Viergever, M. C. A. Dijke, and J. H. Peters, "Reversible intraframe image compression of medical images," *IEEE Trans. Med. Imag.*, vol. 7, pp. 328–336, 1988.
[8] N. V. Boulgouris and M. G. Strintzis, "Orientation-sensitive interpolative pyramids for lossless and progressive image coding," *IEEE Trans. Image Processing*, vol. 9, pp. 710–715, Apr. 2000.
[9] ——, "Optimal progressive lossless image coding using reduced pyramids with variable decimation ratios," *IEEE Trans. Image Processing*, vol. 9, pp. 2117–2123, Dec., 2000.
[10] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmon. Anal.*, vol. 3, no. 2, pp. 186–200, 1996.
[11] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 245–267, 1998.
[12] R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Appl. Comput. Harmon. Anal.*, vol. 5, no. 3, pp. 332–369, 1998.
[13] J. Kovavčević and W. Sweldens, "Wavelet families of increasing order in arbitrary dimensions," *IEEE Trans. Image Processing*, vol. 9, pp. 480–496, Mar. 2000.
[14] R. Claypoole, G. Davis, W. Sweldens, and R. Baraniuk, "Nonlinear wavelet transform for image coding," in *Asilomar Conf. Signals, Systems, Computers*, 1997.
[15] F. J. Hampson and J.-C. Pesquet, "M-band nonlinear subband decompositions with perfect reconstruction," *IEEE Trans. Image Processing*, vol. 7, pp. 1547–1560, Nov. 1998.
[16] D. Taubman, "Adaptive Non-separable lifting transforms for image compression," in *Proc. Int. Conf. Image Processing*, Kobe, Japan, Oct. 1999.
[17] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, pp. 520–540, June 1987.
[18] C. Chrysafis and A. Ortega, "Efficient context-based entropy coding for lossy wavelet image compression," in *Proc. DCC Data Compression Conf.*, Snowbird, UT, Mar. 1997.
[19] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Proc. DCC Data Compression Conf.*, Snowbird, UT, Mar. 1997.
[20] X. Wu, "Lossless compression of continuous-tone images via context selection, quantization and modeling," *IEEE Trans. Image Processing*, vol. 6, pp. 656–664, May 1997.
[21] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[22] D. Tzovaras, "Techniques for the compression of 2- and 3-dimensional images," Ph.D. dissertation, Univ. Thessaloniki, Thessaloniki, Greece, 1997.
[23] A. Papoulis, *Probability Random Variables and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
[24] M. G. Strintzis, "Optimal biorthogonal wavelet bases for signal representation," *IEEE Trans. Signal Processing*, vol. 44, pp. 1406–1418, June 1996.
[25] E. Viscito and J. P. Allebach, "The analysis and design of multidimensional FIR perfect reconstruction filter banks for arbitrary sampling lattices," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 29–41, Jan. 1991.
[26] R. J. Clarke, *Transform Coding of Images*. New York: Academic, 1984.
[27] P. W. Westerink, J. Biemond, and D. E. Boekee, *Subband Coding of Color Images*. Norwell, MA: Kluwer, 1991.
[28] N. K. Bose and Y. Q. Shi, "2-D Wilson spectral factorization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, Jan. 1988.
[29] M. J. T. Smith and S. L. Eddin, "Analysis/synthesis techniques for subband coding," *IEEE Trans. Signal Processing*, vol. 38, pp. 207–214, Aug. 1990.

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.

[30] P. H. A. Nieminen and Y. Neuvo, "A new class of detail preserving filters for image processing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 74–90, Jan. 1987.
[31] P. Heinonen and Y. Neuvo, "FIR-median hybrid filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 832–838, June 1987.
[32] I. Pitas and A. N. Venetsanopoulos, *Non Linear Digital Filters: Principles and Applications*. Norwell, MA: Kluwer, 1990.
[33] M. G. Strintzis, "Optimal pyramidal and subband decompositions for hierarchical coding of noisy and quantized images," *IEEE Trans. Image Processing*, vol. 7, pp. 155–167, Feb. 1998.
[34] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, pp. 30–44, Apr. 1991.
[35] A. Said and W. A. Pearlman, "An image multiresolution representation for lossless and lossy compression," *IEEE Trans. Image Processing*, vol. 5, pp. 1303–1310, Sept. 1996.
[36] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–210, Apr. 1992.
[37] M. A. Viergever and P. Roos, "Hierarchical interpolation: An efficient method for reversible compression of images," *IEEE Eng. Med. Biol. Mag.*, vol. 12, pp. 48–55, Mar. 1993.
[38] X. Wu, "Context quantization with Fisher discriminant for adaptive wavelet image coding," in *Proc. DCC Data Compression Conf.*, Snowbird, UT, 1999, pp. 102–111.

Nikolaos V. Boulgouris (S'96) was born in Greece in 1975. He received the Diploma degree in electrical engineering from the Electrical and Computer Engineering Department, Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece, in 1997. He is currently pursuing the Ph.D. degree in the Electrical and Computer Engineering Department, AUTH.

He holds research and teaching assistanship positions at AUTH. Since 1997, he has participated in several research projects funded by the European Union and the Greek Secretariat of Research and Technology. His research interests include wavelets, image and video coding, image processing, pattern recognition, and multimedia copyright protection.

Mr. Boulgouris is a Member of the Technical Chamber of Greece.

Dr. Dimitrios Tzovaras received the Diploma degree in electrical engineering and the Ph.D. degree in 2-D and 3-D image compression from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 1992 and 1997, respectively.

He is a Senior Researcher with the Informatics and Telematics Institute. Previously, he was a Leading Researcher on 3-D Imaging at AUTH. His main research interests include image compression, 3-D data processing, virtual reality, medical image communication, 3-D motion estimation, and stereo and multiview image sequence coding. His involvement with those research areas has led to the coauthoring of more than 20 articles in refereed journals and more than 50 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. Since 1992, he has been involved in more than 20 projects in Greece funded by the EC and the Greek Ministry of Research and Technology.

Michael Gerassimos Strintzis (M'70–SM'80) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece in 1967, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1969 and 1970, respectively.

He then joined the Electrical Engineering Department, University of Pittsburgh, Pittsburgh, PA, where he served as Assistant (1970–1976) and Associate (1976–1980) Professor. Since 1980 he is Professor of electrical and computer engineering at the University of Thessaloniki, Thessaloniki, Greece, and, since 1999, Director of the Informatics and Telematics Research Institute, Thessaloniki. His current research interests include 2-D and 3-D image coding, image processing, biomedical signal and image processing, and DVD and Internet data authentication and copy protection.

Dr. Strintzis was awarded one of the Centennial Medals of the IEEE in 1984. Since 1999, he has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 22, 2009 at 02:05 from IEEE Xplore. Restrictions apply.