



# Prediction Capabilities of Boolean and Stack Filters for Lossless Image Compression

DOINA PETRESCU\*  
Nokia Research Center, P.O.Box 100, FIN-33720 Tampere, Finland

doina.petrescu@research.nokia.com

IOAN TĂBUȘ  
Signal Processing Laboratory, Tampere University of Technology, P.O. Box 553, FIN-33101 Tampere, Finland

tabus@cs.tut.fi

MONCEF GABBOUJ  
Signal Processing Laboratory, Tampere University of Technology, P.O. Box 553, FIN-33101 Tampere, Finland

moncef@cs.tut.fi

Received August 19, 1996; Revised July 15, 1997

**Abstract.** This paper proposes optimal Boolean, stack, and FIR-Boolean hybrid filters for realizing the prediction stage in lossless grey-level image compression. New optimal design procedures for Boolean filters are introduced, where the optimality criterion is the Error Entropy (EE). The use of the EE-optimal and MAE-optimal Boolean and stack filters in the sequential prediction structure is considered, under different instances: global-optimal, block-optimal, adaptive-size-block-optimal and multiresolution. An extensive simulation study is carried out for analyzing and comparing the performances of the newly introduced predictors and various other sequential predictors. The EE-optimal Boolean predictors prove to be the most efficient predictors. More refined filtering structures, such as block-optimal or adaptive-size-block-optimal are suitable for the prediction task when the prediction mask ought to be small. The proposed progressive transmission structure based on optimal Boolean prediction is shown to outperform HINT [1] progressive lossless coding scheme.

**Key Words:** lossless image compression, stack filters, Boolean filters, FIR Boolean hybrid predictors

## 1. Introduction

Image compression techniques are unanimously recognized as essential tools for the efficient transmission and archiving of image data. There are currently two large classes of compression methods: reversible (lossless), for which perfect reconstruction of the image is possible; and irreversible (lossy), which focuses on obtaining higher compression rates, but allowing only approximate reconstruction [2], [3]. The bitrates reached in lossless compression are usually half or third of the uncompressed image bitrate [4]. These rates are far from the spectacular compression bitrates attained by lossy methods (tenths to hundredths of the uncompressed image bitrate). However, there are several applications where the complete recovery of the original image is strictly required (e.g. medical or satellite imaging [4], [5], [6], [7]) and thus efficient lossless compression is a highly desired objective.

The “spatial redundancy” is the key property of image data that is exploited differently by various compression methods in order to achieve good compression rates. One way to take advantage of this redundancy is to figure out a *prediction model* very accurate for

\* This work was done while with Signal Processing Laboratory, Tampere University of Technology, P.O.Box 553, SF-33101 Tampere, Finland

the image to be transmitted, to make it known also to the receiver, and then to encode and transmit only the unpredictable information, namely the difference between the real and the predicted pixel values.

For lossless image compression, two of the major types of prediction methods are: *sequential prediction methods* and *multiresolution prediction methods* [5], [6], [7].

The prototype of *sequential prediction method* is the linear prediction using a causal prediction template with three pixels (seven such linear predictors are available in JPEG[8]). However, prediction methods based on (adaptive) nonlinear filters have recently been proposed [9], [10], [11], [12], and shown to improve significantly the performance against JPEG predictors.

Sequential prediction methods are restricted by causal constraints to use only partially the two-dimensional spatial information. One way to surpass this disadvantage is to use *multiresolution*, or *hierarchical*, or *pyramidal* methods where each level of the hierarchy represents the entire image, at different spatial resolutions [1], [13]. The lowest resolution level is encoded using a sequential prediction method (like DPCM); the prediction at higher resolutions will be very accurate, due to the use of 360 degree information available for the current pixel from the complete and perfect reconstruction of lower resolution image.

These multiresolution prediction methods are used in progressive image transmission, mainly when lossless transmission must be available only as an option.

This paper will derive optimal nonlinear predictors (Boolean, stack and FIR-Boolean hybrid filters) for the prediction stage in both sequential and multiresolution prediction approaches.

Boolean and stack filters are large classes of nonlinear filters possessing the threshold decomposition property [14]. The optimal design for these filter classes has been developed for different scenarios: using a training framework under the Sum of Absolute Errors criterion [15], [16], [17]; in a model based approach, under the Mean Absolute Error (MAE) criterion [18], [19]; and subject to additional structural constraints [20], [21]. We introduce here the Error Entropy optimal design and build our new procedures on the use of the fast procedures presented in [17].

Boolean and stack filters excel in solving image processing tasks (e.g. restoration, edge detection [22], skeletonization [23]) when the available images are corrupted by impulsive noise. These applications exploit the ability of Boolean filtering structures to cope simultaneously with noise attenuation and detail preservation.

In this paper, we introduce a different class of applications for Boolean filters, namely prediction or interpolation when the available data may be perfectly clean. Other attractive features of this class of filters for small window sizes are thus exploited here: (a) the procedure for the optimal design is simpler than the procedure for linear filter design; (b) the code-length necessary for transmitting the parameters of the filter is smaller for Boolean and stack filters than for linear filters. Due to these properties we will be able to fit very well the predictors to the images to be transmitted, such as to reduce significantly the entropy of the residuals, while the cost incurred by the transmission of the predictor information will not be so high, and globally we shall obtain a lower bitrate than in the fixed predictor case.

Slight variations of the Boolean filter structure will also be considered, in order to customize it for the coding application. Namely, we will consider hybrid structures [24], where

the first level comprises some fixed parameter linear combiners, whose outputs feed the second level, consisting in a Boolean or stack filter. The optimization process will be carried out only for the second level.

The paper is organized as follows. In Section 2 we present new *sequential prediction methods* using optimal nonlinear filter design. In Section 2.1 we present the use of Boolean filters in the prediction stage of lossless compression schemes. Section 2.2 reviews the MAE-optimal design of Boolean and stack filters. In Section 2.3 we introduce the optimal design under EE criterion. All along Section 2 we consider only *global optimal prediction*, where one optimally designed filter is used for the prediction. In Section 3, we study *block-optimal prediction*, where the image is subdivided into small blocks, and (possibly) different predictors are optimally designed for each block. A quadtree procedure is proposed for adaptively selecting the size of the image blocks such as to reduce the cumulative bitrate criterion (this criterion penalizes the residual image entropy as well as the additional bitrate necessary for encoding the predictor and the quadtree itself). Section 4 is devoted to multiresolution prediction and presents the use of optimally designed nonlinear filters for the task of interpolation in a hierarchical multiresolution representation. Experimental results are included in Section 5, providing evidence of the good prediction performances obtained with the new optimal predictors introduced in this paper.

## 2. Sequential Prediction Based on Optimal Boolean and Stack Filtering

### 2.1. Boolean and Stack Filters Used as Predictors

Consider an  $M$ -valued image with  $n_r$  rows and  $n_c$  columns, the current pixel being denoted  $D(i, j)$ :

$$D(i, j) \in \{0, 1, \dots, M-1\}, \quad i = 1, \dots, n_r, \quad j = 1, \dots, n_c$$

The neighborhood  $\mathcal{N}_{D(i,j)}$  will select  $N$  pixels forming the “context” for predicting  $D(i, j)$ . We will arrange the pixels in the neighborhood  $\mathcal{N}_{D(i,j)}$  into an  $N$ -dimensional vector, denoted  $\mathbf{D}(i, j)$ . We require the neighborhood  $\mathcal{N}_{D(i,j)}$  to be causal with respect to the given scanning order, e.g. the neighborhood in Figure 2 is causal when selecting the scanning from left to right and from top to bottom. We propose to use a Boolean filter [17], [25] for predicting the value of  $D(i, j)$ , using the information from  $\mathbf{D}(i, j)$ :

$$\hat{D}(i, j) = \sum_{m=1}^{M-1} f(T_m(\mathbf{D}(i, j))), \quad i = 1, \dots, n_r, \quad j = 1, \dots, n_c \quad (1)$$

where  $T_m$  is the thresholding operator

$$T_m(x) = \begin{cases} 0 & \text{if } x < m \\ 1 & \text{if } x \geq m \end{cases} \quad (2)$$

(which will be applied componentwise to a vector) and  $f$  is a Boolean function. Note that (1) is used here to define Boolean filtering, but there are other efficient ways for the implementation of Boolean filters, avoiding the threshold decomposition, see e.g. [17], [25].

When the Boolean function  $f$  is positive, i.e. it can be represented in a sum-of-product form containing only uncomplemented variables, the filter is a stack filter [14] and the output can be computed in the integer domain using min and max operators (acting on the pixels in  $\mathcal{N}_{D(i,j)}$ ). More details about the properties of stack and Boolean filters can be found in [17], [25], [26], [27]. In image restoration applications, stack filters have been found more robust than Boolean filters [17]. However, the optimal design procedure for stack filters requires extra-computational complexity for enforcing the positivity of the Boolean function [15], [17], [18].

### 2.1.1. FIR-Boolean Hybrid Filters

We introduce a new filtering architecture for prediction, combining linear predictors with optimal Boolean filters, aiming to achieve higher prediction performance than each one separately. The FIR-Boolean hybrid (FBH) filtering architecture presented in Figure 1 consists in two filtering levels [28]. At each pixel position, the vector  $\mathbf{X}$  containing the values of the pixels inside the processing mask is processed in parallel in the first level by several FIR filters. The outputs of the linear filters are arranged in a vector and fed into the second level, consisting in a Boolean filter. The output of the FIR-Boolean hybrid filter, having  $N$ -pixel processing mask, and  $Q$  FIR filters on the first level is then computed as:

$$\begin{aligned} Y_{FBH(N,Q)} &= \sum_{m=1}^{M-1} f(T_m([Y_1, \dots, Y_Q])) \\ &= \sum_{m=1}^{M-1} f\left(T_m\left(\left[\sum_{i=1}^N a_{1,i} X_i, \dots, \sum_{i=1}^N a_{Q,i} X_i\right]\right)\right), \end{aligned} \quad (3)$$

where  $\{a_{j,i}\}_{i=1}^N$  are the coefficients of the filter  $\text{FIR}_j$ ,  $j = 1, \dots, Q$ .

For the prediction task we used FIR-Boolean hybrid filters with fixed linear predictors on the first level of the architecture and an optimal Boolean filter on the second level.

## 2.2. MAE-Optimal Boolean and Stack Filters

In general Boolean filters exhibit very different prediction performances, and one way to assess their predicting performance is to use the Sum of Absolute Errors criterion (that is sometimes reported as the Mean Absolute Error (MAE) criterion)

$$J_f = \frac{1}{T} \sum_{i=1}^{n_r} \sum_{j=1}^{n_c} |D(i, j) - \hat{D}(i, j)| = \frac{1}{T} \sum_{i=1}^{n_r} \sum_{j=1}^{n_c} \left| D(i, j) - \sum_{m=1}^{M-1} f(T_m(\mathbf{D}(i, j))) \right| \quad (4)$$

In the case of stack filters, the positive Boolean function minimizing the prediction criterion (4) can be found by solving a linear programming problem [18] using e.g. the fast algorithms provided in [15], [17]. On the other hand, for the Boolean filter class, the optimal predictor minimizing criterion (4) can be extremely easily found for the binary case

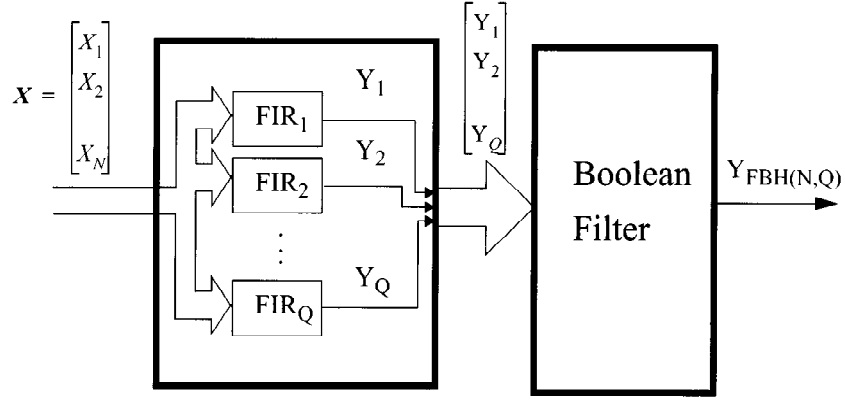


Figure 1. FIR-Boolean filtering architecture.

( $M = 1$ ). For  $M > 1$ , the problem of finding the global optimal solution has an exponential complexity; however it is easy to find a suboptimal solution w.r.t. criterion (4). This suboptimal solution will minimize an upper bound of  $J_f$ , i.e. it will globally minimize the criterion

$$J_f^b = \frac{1}{T} \sum_{i=1}^{n_r} \sum_{j=1}^{n_c} \sum_{m=1}^{M-1} |T_m(D(i, j)) - T_m(\hat{D}(i, j))| > J_f \quad (5)$$

For each binary context  $\mathbf{v}_\ell$ ,  $\ell = 1, \dots, 2^N$ , the suboptimal design procedure will first collect the counts

$$N_0(\mathbf{v}_\ell) = \text{Card}\{(m, i, j) | T_m(\mathbf{D}(i, j)) = \mathbf{v}_\ell, T_m(D(i, j)) = 0\} \quad (6)$$

$$N_1(\mathbf{v}_\ell) = \text{Card}\{(m, i, j) | T_m(\mathbf{D}(i, j)) = \mathbf{v}_\ell, T_m(D(i, j)) = 1\} \quad (7)$$

which can be used to express the criterion (5) as

$$J_f^b = C_0 + \sum_{\ell=1}^{2^N} c(\mathbf{v}_\ell) f(\mathbf{v}_\ell) \quad (8)$$

where

$$c(\mathbf{v}_\ell) = N_0(\mathbf{v}_\ell) - N_1(\mathbf{v}_\ell) \quad (9)$$

$$C_0 = \sum_{\ell=1}^{2^N} N_1(\mathbf{v}_\ell) \quad (10)$$

The value of the Boolean function for binary context  $\mathbf{v}_\ell$ ,  $\ell = 1, \dots, 2^N$  is assigned as

$$f(\mathbf{v}_\ell) = \begin{cases} 0 & \text{if } c(\mathbf{v}_\ell) > 0 \\ 1 & \text{if } c(\mathbf{v}_\ell) \leq 0 \end{cases} \quad (11)$$

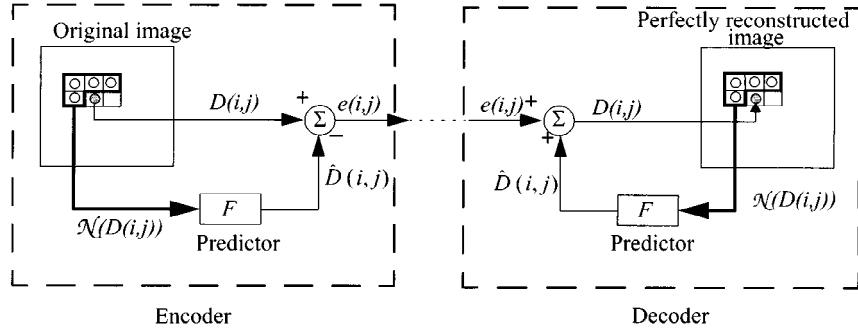


Figure 2. Lossless coding of images using sequential prediction technique.

When  $c(\mathbf{v}_\ell) = 0$ , either 0 or 1 can be assigned in (11), with no difference in the prediction performance.

In our experiments, we will also use FIR-Boolean hybrid (FBH) predictors, which include in the processing mask of the Boolean filter not only the pixels from the prediction mask, but also some linear combinations of them. The optimization will however be performed only for the Boolean function, the weights of the linear combinations being kept fixed.

### 2.3. Optimal Boolean and Stack Filters Under the Entropy of Errors Criterion

#### 2.3.1. Modeling the Entropy of Prediction Errors

A good evaluation of the lower bound on the average number of bits needed to encode the prediction errors

$$e_f(i, j) = D(i, j) - \sum_{m=1}^{M-1} f(T_m(\mathbf{D}(i, j))) \quad (12)$$

is the entropy

$$H(p_{e_f}) = - \sum_{e=-(M-1)}^{M-1} p_{e_f}(e) \log_2 p_{e_f}(e). \quad (13)$$

where  $p_{e_f}(e)$  is the relative frequency of occurrence of the prediction error value  $e$ ,  $e \in \{-(M-1), -(M-2), \dots, 0, \dots, (M-1)\}$ . Denoting by  $h_f(e)$  the histogram of the prediction errors of an image, where

$$h_f(e) = \text{Card}\{(i, j) | e_f(i, j) = e\} \quad (14)$$

and the total number of pixels by  $T = n_r n_c$ , the relative frequency of error occurrence will be  $p_{e_f}(e) = h_f(e)/T$  and hence

$$H(h_f) = \log_2 T - \frac{1}{T} \sum_{e=-(M-1)}^{M-1} h_f(e) \log_2 h_f(e) \quad (15)$$

$$= \log_2 T - \frac{1}{T} \sum_{i=1}^{n_r} \sum_{j=1}^{n_c} \log_2 h_f(e_f(i, j)) \quad (16)$$

where (16) is simply rephrasing (15), but with a more useful interpretation, as the sample mean over the image of the quantity  $\log_2 h_f(e_f(i, j))$ . In order to introduce the modeling environment for computing the entropy we will make some considerations regarding the histogram  $h_f(e_f(i, j))$ .

The histograms of natural images manifest a high variability (various number of local maxima, types of monotonicity between local minima and maxima: ramp, exponential, etc.) [29]. Totally different patterns are encountered with the prediction error histograms: they are highly regular for the range of medium and small errors; they present a symmetrical lobe, centered at  $e = 0$  and have monotonic sides, the monotonicity being piecewise exponential. The very large errors occur only seldom and no pattern can be fit on this range of the histogram (see Figure 3). Therefore, we will model the histogram as a piecewise symmetrical doubly exponential function

$$\hat{h}_\Theta(e) = \begin{cases} \mu_1 2^{-\lambda_1 |e|}, & \text{for } \theta_0 = 0 \leq |e| < \theta_1 \\ \mu_2 2^{-\lambda_2 |e|}, & \text{for } \theta_1 \leq |e| < \theta_2 \\ \dots & \dots \\ \mu_p 2^{-\lambda_p |e|}, & \text{for } \theta_{p-1} \leq |e| < \theta_p = M \end{cases} \quad (17)$$

where  $\Theta$  groups all parameters of (17),  $\Theta = \{\theta_1, \dots, \theta_{p-1}, \mu_1, \dots, \mu_p, \lambda_1, \dots, \lambda_p\}$ . The case  $p = 1$  corresponds to the Laplacian probability distribution of the errors and has been long advocated when analyzing the errors in DPCM structures [2]. The fitting of  $\hat{h}_\Theta(e)$  to an experimental histogram,  $h_f(e)$ , can be accomplished by a LS procedure minimizing the criterion

$$\begin{aligned} J_{h_f}(\Theta) &= \sum_{|e|=0}^{M-1} \left( \log_2(h_f(e)) - \log_2(\hat{h}_\Theta(e)) \right)^2 \\ &= \sum_{|e|=0}^{\theta_1-1} (\log_2(h_f(e)) - \log_2 \mu_1 + \lambda_1 |e|)^2 + \dots \\ &\quad + \sum_{|e|=\theta_{p-1}}^{\theta_p-1} (\log_2(h_f(e)) - \log_2 \mu_p + \lambda_p |e|)^2 \end{aligned} \quad (18)$$

where the number of segments,  $p$ , and the segment boundaries,  $\theta_1 \dots, \theta_{p-1}$ , are selected by the user (e.g. by inspection of the curve  $h_f$ ), while the parameters  $\mu_1, \dots, \mu_p$  and  $\lambda_1, \dots, \lambda_p$  will immediately result by solving the least square estimation problem (18) with  $2p$  parameters (which in turn, can be decoupled to minimizing  $p$  LS criteria (18), each depending on two parameters).

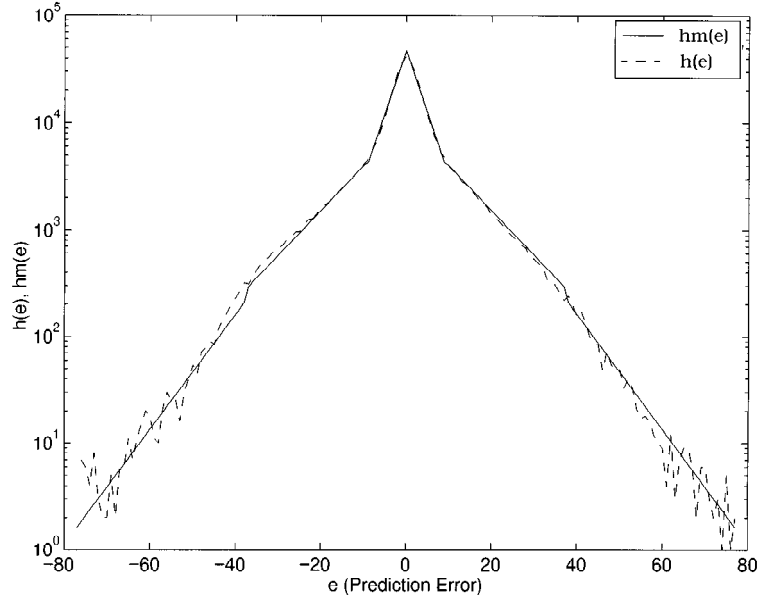


Figure 3. Modeling the histogram of prediction errors using piecewise exponential functions. (--)  $h(e)$  is the experimental histogram (image *Barb1*, Boolean predictor, with 10-pixel mask); (—)  $h_m(e)$  is a fitted model to the histogram, with  $k = 3$  exponentials:  $\lambda_1 = 0.2680$ ,  $\lambda_2 = 0.0968$ ,  $\lambda_3 = 0.1245$ ;

In all the experiments reported in this paper, we restricted to  $p = 2$  and have arbitrarily selected  $\theta_1 = 9$  as the threshold between the low and high error models. Intuitively, we will treat differently the prediction in smooth areas, where the errors are very low, and the prediction in the areas with many details, where the errors given by our causal predictors will inherently be high (no matter what predictor we use). The compound two-exponential-rates model of the entropy will establish how to compromise the prediction accuracy in small error areas and large error areas, such as to minimize the entropy of the whole error image.

### 2.3.2. The EE-optimal Boolean prediction problem

We will denote by  $\Theta_f$  the optimal parameters minimizing (18) for the experimental histogram  $h_f$ . Denote

$$\mathcal{E}_k(f) = \{(i, j) | \theta_{k-1} \leq |e_f(i, j)| < \theta_k\} \quad (19)$$

the set of pixel locations resulting when modeling the histogram  $h_f$  with the model in (17) for  $p = 2$ . Now the entropy of the prediction error (16) can be written as

$$H(h_f) = \log_2 T - \frac{1}{T} \sum_{i=1}^{n_r} \sum_{j=1}^{n_c} \log_2 h_f(e_f(i, j)) \quad (20)$$

$$\begin{aligned} &\approx \log_2 T - \frac{1}{T} \sum_{i=1}^{n_r} \sum_{j=1}^{n_c} \log_2 \hat{h}_{\Theta_f}(e_f(i, j)) \\ &= \log_2 T - \frac{1}{T} \left\{ \sum_{(i,j) \in \mathcal{E}_1(f)} (\log_2 \mu_{1f} - \lambda_{1f} |e_f(i, j)|) \right. \\ &\quad \left. + \sum_{(i,j) \in \mathcal{E}_2(f)} (\log_2 \mu_{2f} - \lambda_{2f} |e_f(i, j)|) \right\} \quad (21) \end{aligned}$$

$$= \mu_{0f} + \frac{1}{T} \left\{ \sum_{(i,j) \in \mathcal{E}_1(f)} \lambda_{1f} |e_f(i, j)| + \sum_{(i,j) \in \mathcal{E}_2(f)} \lambda_{2f} |e_f(i, j)| \right\} \quad (22)$$

where

$$\mu_{0f} = \log_2 T - \frac{1}{T} \text{Card}(\mathcal{E}_1(f)) \log_2 \mu_{1f} - \frac{1}{T} \text{Card}(\mathcal{E}_2(f)) \log_2 \mu_{2f} \quad (23)$$

**The connection between MAE criterion and EE criterion** Observe that in the case  $p = 1$ , of modeling the histogram using only one exponential function (or, equivalent, using  $p$  exponentials with the same rates  $\lambda_1 = \dots = \lambda_p$ ) the entropy reduces to

$$H(h_f) \approx \mu_{0f} + \frac{\lambda_{1f}}{T} \sum_{(i,j)} |e_f(i, j)| \quad (24)$$

In our experiments, we found that the parameters  $\mu_{0f}$  and  $\lambda_{1f}$  do not depend significantly on  $f$ , in a neighborhood around the filter minimizing  $\sum_{(i,j)} |e_f(i, j)|$ , i.e. the entropy will depend almost linearly on the sum of absolute errors criterion. This explains why generally a low entropy of error is obtained with MAE-optimal designed Boolean predictors, optimal when fitting a  $p = 1$  model to the experimental histogram.

**The connection between weighted MAE criterion and EE criterion** When using the histogram model with  $p = 2$ , improvements over the case  $p = 1$  are expected. The problem of designing a predictor optimal w.r.t. the entropy of the errors, can be stated as to find the Boolean function  $f^*$  which minimizes the criterion

$$H(h_f) \approx \mu_{0f} + \frac{1}{T} \left\{ \sum_{(i,j) \in \mathcal{E}_1(f)} \lambda_{1f} |e_f(i, j)| + \sum_{(i,j) \in \mathcal{E}_2(f)} \lambda_{2f} |e_f(i, j)| \right\} \quad (25)$$

Note that the criterion

$$H_{WMAE}(f) = \frac{1}{T} \left\{ \sum_{(i,j) \in \mathcal{E}_1} \lambda_1 |e_f(i,j)| + \sum_{(i,j) \in \mathcal{E}_2} \lambda_2 |e_f(i,j)| \right\} \quad (26)$$

is a weighted MAE criterion for some given sets  $\mathcal{E}_1, \mathcal{E}_2$  and constants  $\lambda_1, \lambda_2$  and obviously is much simpler to minimize than (25).

**Exact solution of the minimization problem** We introduce the EE-optimal Boolean predictor as the predictor minimizing the highly nonlinear criterion (20). However the exact solution to the minimization problem involves extensive search over the set of all  $2^{2^N}$  predictors. The evaluation of the criterion (20) for each filter is clearly untractable for common values of the prediction mask size ( $N \approx 10$ ).

**Bootstrap procedure** The first practical attempt to find a suboptimal solution is to transform the minimization of the nonlinear criterion into a sequence of weighted MAE criteria minimizations, by resorting to the bootstrap technique, where during one minimization stage, the values found in the previous stage,  $\mathcal{E}_1(f^{k-1}), \mathcal{E}_2(f^{k-1}), \lambda_1(f^{k-1}), \lambda_2(f^{k-1})$ , are used instead of the true ones.

This will result in an iterative two-step procedure for finding  $f^*$ . In the first step, we solve for the  $f^k$  that minimizes the criterion

$$H_{WMAE}(f^k) = \frac{1}{T} \left\{ \sum_{(i,j) \in \mathcal{E}_1(f^{k-1})} \lambda_{1f^{k-1}} |e_{f^k}(i,j)| + \sum_{(i,j) \in \mathcal{E}_2(f^{k-1})} \lambda_{2f^{k-1}} |e_{f^k}(i,j)| \right\} \quad (27)$$

which, being a weighted MAE criterion, can be minimized using a procedure very similar to that minimizing MAE criterion.

In the second step (after  $f^k$  was found), we partition the pixels of the image into the sets  $\mathcal{E}_1(f^k), \mathcal{E}_2(f^k)$ , then compute the histogram  $h_{f^k}$  of the prediction errors and finally the model  $\Theta_{f^k}$  is fitted to the histogram, making now available all necessary information for the new iteration,  $k + 1$ . The recursion is initialized at  $k = 0$  with  $f^0$ , the predictor minimizing the criterion  $J_f^b$  (5).

This procedure is iterated until no more improvements are found with respect to the previous iteration. We found in our experiments that this procedure will not always evolve properly, toward a predictor with entropy better than the MAE-optimal filter, often getting stuck in local minima of the criterion (25).

Another drawback of this iterative two-step search is the high complexity, involving at each iteration one pass through the image, collecting the histogram of the errors, modeling the histogram to obtain the new parameters  $\Theta$  and finally minimizing (27) corresponding to the new values of the parameters  $\lambda_1$  and  $\lambda_2$ .

**EE-(sub)optimal design procedure** The second and the most effective technique for minimizing the criterion (20) is to freeze the sets  $\mathcal{E}_1(f), \mathcal{E}_2(f)$  to the sets  $\mathcal{E}_1(f^0), \mathcal{E}_2(f^0)$ , corresponding to the MAE-optimal solution,  $f^0$ , and instead of solving the modeling problems which find  $\lambda_1(f^{k-1}), \lambda_2(f^{k-1})$ , to only check the filters minimizing the criterion (26)

corresponding to any values of the pairs  $(\lambda_1, \lambda_2)$ . The problem is thus reduced to carefully planing the search, over a reduced set of functions  $f$ , those who minimize the criterion (26) for different values of  $\lambda_1$  and  $\lambda_2$  and since only a small number of such optimal  $f$  exists, this technique will prove more efficient than exhaustive search.

Moreover, since criterion (26) was obtained by a series of approximations of criterion (20), the final solution will be selected from the restrained search space of functions  $f$  and it will be the one minimizing the criterion (20).

First we concentrate on finding the restrained set of functions  $f$ , which are minimum points of the criteria (26), for all values of  $\lambda_1$  and  $\lambda_2$ . The criterion (26) has the same mathematical form as the criterion found in the optimization of nonlinear filters with structural constraints [30] and we customize in the following the procedure given in [30] for the prediction setting. The criterion to be minimized in the predictor design procedure is the following upper bound of (26)

$$J^b(f) = \frac{\lambda_1}{T} \sum_{(i,j) \in \mathcal{E}_1} \sum_{m=1}^{M-1} |T_m(D(i,j)) - f(T_m(\mathbf{D}(i,j)))| + \frac{\lambda_2}{T} \sum_{(i,j) \in \mathcal{E}_2} \sum_{m=1}^{M-1} |T_m(D(i,j)) - f(T_m(\mathbf{D}(i,j)))| \quad (28)$$

We introduce the cost coefficients (9) for the binary contexts  $\mathbf{v}_\ell \in \{0, 1\}^N$  for the two sets of pixels:

$$c^{[\mathcal{E}_k]}(\mathbf{v}_\ell) = N_0^{[\mathcal{E}_k]}(\mathbf{v}_\ell) - N_1^{[\mathcal{E}_k]}(\mathbf{v}_\ell), \quad k = 1, 2 \quad (29)$$

where  $N_0^{[\mathcal{E}_k]}(\mathbf{v}_\ell)$  and  $N_1^{[\mathcal{E}_k]}(\mathbf{v}_\ell)$  are defined by (6), (7), additionally restricting  $(i, j) \in \mathcal{E}_k$ . Introducing the *mixing cost coefficients* [30], for the binary vectors  $\mathbf{v}_\ell \in \{0, 1\}^N$

$$c^{[mix]}(\mathbf{v}_\ell) = \lambda_1 c^{[\mathcal{E}_1]}(\mathbf{v}_\ell) + \lambda_2 c^{[\mathcal{E}_2]}(\mathbf{v}_\ell) \quad (30)$$

the criterion (28) becomes

$$J^b(f) = C_0^{[mix]} + \frac{1}{T} \sum_{\ell=1}^{2^N} c^{[mix]}(\mathbf{v}_\ell) f(\mathbf{v}_\ell) \quad (31)$$

where  $C_0^{[mix]}$  does not depend on  $f$ . The predictor  $f^{\lambda_1, \lambda_2}$  minimizing (31), for fixed  $\lambda_1$  and  $\lambda_2$ , is obtained assigning the values of  $f^{\lambda_1, \lambda_2}(\mathbf{v}_\ell)$  according to the sign of  $c^{[mix]}(\mathbf{v}_\ell)$  as in (11).

**LEMMA** *There are  $K \leq 2^N$  positive numbers,  $v_1, \dots, v_K$  with  $v_0 = 0 < v_1 < \dots < v_K < v_{K+1} = \infty$ , such that to any  $\lambda_1, \lambda_2 > 0$  with  $\frac{\lambda_1}{\lambda_2} \in (v_k, v_{k+1})$ , corresponds the same Boolean function  $f^{\lambda_1, \lambda_2}$  minimizing (31) (in (31) the cost coefficients are given by (30) with fixed pixel sets  $\mathcal{E}_1$  and  $\mathcal{E}_2$ ).*

**Proof:** The Boolean function minimizing (31),  $f^{\lambda_1, \lambda_2}(\mathbf{v}_\ell)$ , is assigned according to the sign of  $c^{[mix]}(\mathbf{v}_\ell) = \lambda_1 c^{[\mathcal{E}_1]}(\mathbf{v}_\ell) + \lambda_2 c^{[\mathcal{E}_2]}(\mathbf{v}_\ell) = \lambda_2 \left( \frac{\lambda_1}{\lambda_2} c^{[\mathcal{E}_1]}(\mathbf{v}_\ell) + c^{[\mathcal{E}_2]}(\mathbf{v}_\ell) \right)$  which in its turn depends only on the ratio  $\nu = \frac{\lambda_1}{\lambda_2}$ . If the value

$$\nu_{change} = -\frac{c^{[\mathcal{E}_2]}(\mathbf{v}_\ell)}{c^{[\mathcal{E}_1]}(\mathbf{v}_\ell)} \quad (32)$$

is positive, then there will be a change in  $f^{\lambda_1, \lambda_2}(\mathbf{v}_\ell)$  only when  $\nu = \frac{\lambda_1}{\lambda_2}$  crosses the value  $\nu_{change}$ . Now we collect all the positive  $\nu_{change}$  corresponding to all  $\mathbf{v}_\ell$ , (at most  $2^N$  numbers) and then order them, obtaining the string  $\nu_0 = 0, \nu_1, \dots, \nu_K, \nu_{K+1} = \infty$ . For any pair  $\lambda_1, \lambda_2$  with  $\frac{\lambda_1}{\lambda_2} \in (\nu_k, \nu_{k+1}), k = 0, 1, \dots, K$ , the predictor  $f^{\lambda_1, \lambda_2}$  minimizing (31) will be the same. ■

The possible solutions  $f^{\lambda_1, \lambda_2}$  will change from the MAE-optimal Boolean predictor for the pixels in the set  $\mathcal{E}_1$  (easily predictable), to the MAE-optimal Boolean predictor for the pixels  $\mathcal{E}_2$  (difficult to predict).

The number  $K + 1$  of different solutions  $f^{\lambda_1, \lambda_2}$  is in practice much lower than  $2^N$ , since  $\nu_{change}$  (32) are usually negative for most of the binary contexts  $\mathbf{v}_\ell$  (in our present prediction experiments we found in average 4 different predictors for  $N = 3$  and about 200 different predictors for  $N = 10$ ). Table 1 summarizes the main steps of the EE-suboptimal design procedure presented in this paragraph. We will denote in the sequel, for short, this procedure by *EE-optimal design*, even if the result of the procedure may be only suboptimal, due to all approximations in the derivation of the algorithm.

However, since the partition of the image into the sets  $\mathcal{E}_1(f^0), \mathcal{E}_2(f^0)$  corresponds to the optimal MAE solution, the resulting solution of our procedure will have the entropy less or equal to the entropy of the MAE-optimal filter, which was experimentally confirmed in all our experiments.

We can check easily the entropy (15) for all those different filters to find the best filter to use as predictor. This search can be made more efficient by observing that the scalar  $\lambda$  parameterizes the filters to be tested, and therefore efficient line search methods can be used in Step 5 (e.g. the golden section method).

### 3. Sequential Prediction Using Block-Optimal Boolean Filters

#### 3.1. Fixed Size Block-Optimal Prediction

Locally adaptive (block-optimal) Boolean filters have been introduced in [31] for image restoration applications, where they were shown to outperform the one-block optimal Boolean filters. Similar improvement will be obtained in the present prediction application, but with the cost of increasing the predictor complexity.

The image is subdivided into small blocks and one MAE-optimal Boolean predictor is fitted to each block. All the predictors must be transmitted altogether with the prediction errors, in order to recover the image at the decoder.

<p><i>Step 1</i> Find the MAE-optimal Boolean predictor <math>f^0</math> by using (6)–(11).</p> <p><i>Step 2</i> Find the sets <math>\mathcal{E}_1</math> (easily predictable pixels) and <math>\mathcal{E}_2</math> (difficult to predict pixels), for the predictor <math>f^0</math> by using (19).</p> <p><i>Step 3</i> Compute the cost coefficients <math>\{c^{[\mathcal{E}_1]}(\mathbf{v}_\ell)\}_{\ell=1}^{2^N}</math>, <math>\{c^{[\mathcal{E}_2]}(\mathbf{v}_\ell)\}_{\ell=1}^{2^N}</math> by using (29).</p> <p><i>Step 4</i> Find the values of <math>v = \frac{\lambda_1}{\lambda_2} \in [0, \infty)</math> where <math>f^{\lambda_1, \lambda_2}</math> may change:</p> <p>For each binary context <math>\mathbf{v}_\ell</math> if <math>c^{[\mathcal{E}_1]}(\mathbf{v}_\ell)</math> and <math>c^{[\mathcal{E}_2]}(\mathbf{v}_\ell)</math> have different signs, then compute</p> $v_{change} = -\frac{c^{[\mathcal{E}_2]}(\mathbf{v}_\ell)}{c^{[\mathcal{E}_1]}(\mathbf{v}_\ell)} \quad (33)$ <p>Order increasingly the values <math>v_{change}</math> obtaining the sequence <math>v_1, \dots, v_K</math>. Assign <math>v_0 = 0</math>, <math>v_{K+1} = v_K + 1</math>.</p> <p><i>Step 5</i> For each <math>k = 1, \dots, K + 1</math> compute</p> <ul style="list-style-type: none"> <li>• the mixing factors <math>\lambda_2 = 1</math> and <math>\lambda_1 = (v_{k-1} + v_k)/2</math>.</li> <li>• the set of mixed cost coefficients <math>\{c^{[mix]}(\mathbf{v}_\ell)\}_{\ell=1}^{2^N}</math> by using (30).</li> <li>• assign the Boolean function</li> </ul> $f^{\lambda_1, \lambda_2}(\mathbf{v}_\ell) = \begin{cases} 0 & c^{[mix]}(\mathbf{v}_\ell) > 0 \\ 1 & c^{[mix]}(\mathbf{v}_\ell) \leq 0 \end{cases} \quad (34)$ <ul style="list-style-type: none"> <li>• compute the entropy of errors for the predictor <math>f^{\lambda_1, \lambda_2}</math> by using (14), (15).</li> </ul> <p><i>Step 6</i> Select the best (w.r.t. (15)) predictor obtained in the search at Step 5.</p>
--

Table 1. EE-(sub)optimal design procedure.

The prediction efficiency is measured using a cumulative global criterion,  $\mathcal{R}$ , which includes both the entropy of the residual image and the additional bitrate needed to encode each local predictor.

For the case of regular partition of the image into  $n_b$  equally sized blocks the cumulative criterion must be computed as follows

$$\mathcal{R} = H + 2^N \cdot n_b / T \quad (35)$$

where  $H$  is the entropy of the errors (computed using the histogram of the *whole error image*),  $N$  denotes the prediction window size and  $2^N$  is the number of bits necessary for coding each predictor.

The regular block decomposition of the image is conceptually and procedurally simple, but may be inefficient; e.g. for flat areas different filters for each block will not improve too much the prediction performance, while a high cost would have to be paid for transmitting each filter; while for areas with a high density of details, the splitting into very low size blocks might improve significantly the overall criterion  $\mathcal{R}$ .

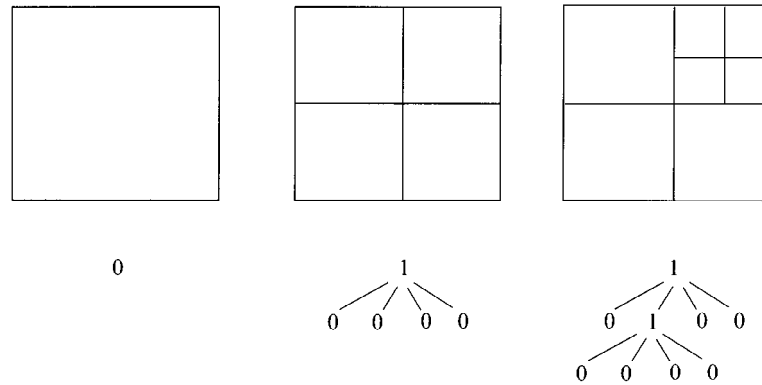


Figure 4. Quad tree codes for three types of blocks.

The next subsection presents a quadtree procedure used for adaptively setting the size in the image partition.

### 3.2. Adaptive Size Block-Optimal Prediction

Quadtrees have been widely used in connection with image compression techniques, see e.g. [32]. We use in this paper a top-down strategy to perform adaptive block size partition. We start with one large block and we check whether subdividing it into four equally sized subblocks in order to design an optimal filter for each of them separately would yield improvement in performance (by reducing the global criterion  $\mathcal{R}$ ). The highest weight in the criterion  $\mathcal{R}$  is given by the entropy of prediction errors for the whole image,  $H$ . This will make the decision on further subdividing each block highly dependent on how the rest of the image is already partitioned, and hence this decision may prove only suboptimal.

Reiterating the top-down splitting decisions starting from the final partition obtained in the first top-down pass may provide better decisions, but with an increase in computational effort.

We start the procedure using an initial partition of the image into blocks of equal (medium) size,  $L \times L$ , ( $32 \times 32$  in our experiments), for which the global criterion  $\mathcal{R}$  is computed from (35). The image will be scanned, block by block. A recursive procedure decides whether or not to split the current block and one bit,  $b$ , is allocated to encode the decision for each subblock. A quadtree is built and coded for each  $L \times L$  block, as shown in Figure 4. The  $L \times L$  blocks are scanned in left-right, top-bottom order, and the quadtree is coded in top-down mode. For the blocks shown in Figure 4 the coding stream will be

0 1 0 0 0 0 1 0 1 0 0 0 0 0 0

**Quadtree procedure for subdividing images for prediction** We give a summary description of the recursive splitting procedure. Denote by  $\mathcal{B}$  the block which is currently being checked whether to split it or not. Denote by  $h_{out}(e)$  the histogram of the errors at pixel locations outside block  $\mathcal{B}$ .

1. Compute the optimal predictor  $f_{\mathcal{B}}$  for the overall block and  $f_{\mathcal{B}_1}, \dots, f_{\mathcal{B}_4}$ , the optimal predictors for the subblocks  $\mathcal{B}_1, \dots, \mathcal{B}_4$ .
2. Compute the prediction errors and the histograms  $h_{\mathcal{B}}(e), h_{\mathcal{B}_1}(e), \dots, h_{\mathcal{B}_4}(e)$  and the entropy of these errors for one block,  $H_{no-split} = H(h_{out}(e) + h_{\mathcal{B}}(e))$ , and for the four blocks  $H_{split} = H(h_{out}(e) + h_{\mathcal{B}_1}(e) + \dots + h_{\mathcal{B}_4}(e))$ .
3. Decide to split the block if

$$H_{no-split} > H_{split} + \frac{1}{T}(3 \times 2^N + 4) \tag{36}$$

$3 \times 2^N$  is the cost for extra-coding the predictors for the four subblocks and that for transmitting the code for the quadtree branching is 4 bits.

4. If the block is split, and if it is not at the deepest level accepted for splitting ( $4 \times 4$  in our experiments), call the present procedure for each of the resulting subblocks.

#### 4. Optimal Boolean Filters Used for Multiresolution Prediction

In sequential prediction, the scanning order of the pixels in the image is kept the same for the whole image, making possible the recovery of the pixels exactly in the scanning order. Hierarchical prediction comprises a number of different stages of sequential prediction where the original image is subsampled at different resolutions. Various multiresolution prediction structures using linear predictors have been proposed [1], [13], [33]. A multiresolution prediction structure starting with the original image subsampled horizontally and vertically by a factor of 2 is presented in Figure 5. The principle is similar when starting with lower resolutions (subsampling by factors of 8, 4 or 2 have been commonly used studied for various predictors). Each stage has its own scanning order, prediction mask and predictor.

In the first stage, the pixels  $(2 \times i + 1, 2 \times j + 1)$  (circles in Figure 5) are sequentially predicted, using a causal prediction mask (which involves only pixels at locations  $(2 \times k + 1, 2 \times l + 1)$ , where  $i, k \in \{0, \dots, \lfloor (n_r - 1)/2 \rfloor\}$  and  $j, l \in \{0, \dots, \lfloor (n_c - 1)/2 \rfloor\}$ ). This stage achieves the transmission of the subsampled version of the original image to the decoder. Stage II will continue with the sequential prediction of the pixels at locations  $(2 \times i, 2 \times j)$ , with  $i$  and  $j$  as above (triangles in Figure 5). The prediction mask may now contain any combination of pixels (not restricted by causality) from the image transmitted in Stage I and from the pixels already available in the sequential Stage II. The prediction accuracy is significantly increased at this stage, due to the availability of pixels in the subsampled image (the term interpolation is often used instead of prediction, due to the apparent noncausality

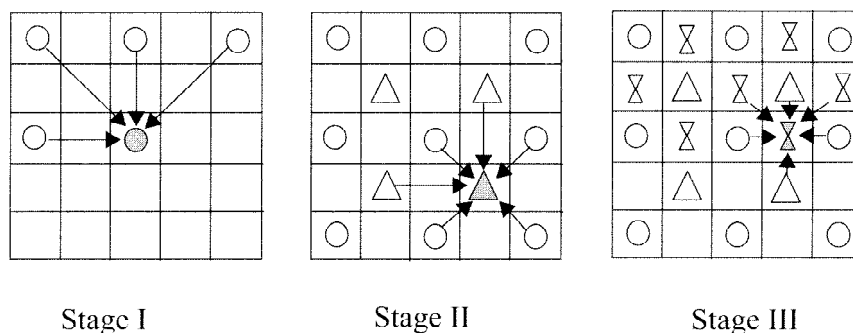


Figure 5. Predicting at different resolution levels.

of the prediction mask). Half of the image pixels are available at the end of the second stage of sequential prediction.

For the remaining pixels, predicted in Stage III, the information available is even richer than in the second stage. Six pixels out of eight in 8-connectivity are available and therefore the prediction will have a very high accuracy.

For each stage of this multiresolution structure there will be a different prediction mask. The best Boolean filter will be fitted in each stage, using any of the sequential prediction procedures developed in Sections 2 and 3.

Different interpolation techniques based on median filters and FIR median hybrid filters have been proposed [34], [35] and shown to perform better than linear interpolative techniques. We will analyze the performances of MAE-optimal Boolean and FIR-Boolean hybrid filters for multiresolution prediction and compare them with previously reported results.

## 5. Experimental Results

### 5.1. General Setting of the Experiments

The experiments reported in this paper use eight  $576 \times 720$  images (Balloon, Barb1, Barb2, Boats, Girl, Gold, Hotel and Zelda) from the JPEG test image set (images used during the development of JPEG standard). Only the Y (luminance) component, represented with 8-bits/pixel, of each image was used in our tests.

In this section, we concentrate on reporting results pertaining to the prediction capabilities of different classes of Boolean filters. Therefore, we will report the entropy  $H$  of the errors (given by (13), (15) or (16)), which is a lower bound on the required codelength per pixel obtained with independent source coding techniques (Huffman, arithmetic coding). We prefer to take as measure of performance the entropy of the errors and not the codelength required for encoding the images since the encoding stage can be performed using a large

variety of algorithms, with various performances. However, the codelength needed to encode the predictors themselves will be taken into account, when this codelength becomes significant (note that when using one single optimal predictor, say of size  $N = 10$ , for the overall image, the codelength for encoding the predictor will be only 0.0025 bits per pixel).

For comparison purposes, we include here also the performances of the standardized JPEG predictors and of other predictors recently reported, like the median adaptive predictor ([11]) and the gradient adjusted predictor used by the CALIC system ([36]).

### 5.2. Sequential Prediction

We performed six experiments in order to analyze various alternatives in setting the type of predictors (Boolean, stack, FIR-Boolean hybrid filters), the structural parameters (size of prediction mask) and the type of prediction (global, local, or locally adaptive).

**Experiment #1. Different predictors with 3 pixels prediction mask.** In this experiment the performances of Boolean, stack and FIR-Boolean hybrid predictors are examined, when the prediction mask is 3 pixels wide:

$$\begin{matrix} \circ & \circ \\ \circ & \bullet \end{matrix} \tag{37}$$

where the black circle represents the predicted pixel.

The values of the error entropy for different linear and nonlinear predictors are compared in Table 2. The entropy of the original image is reported in the second column. The JPEG column corresponds to the JPEG predictor having the best average of error entropy for the 8 images. The Boolean(3) and Stack(3) columns correspond to EE-optimal Boolean and MAE-optimal stack predictors. The last two columns correspond to the Martucci's MedAP, median adaptive predictor, which certainly is a FBH(3,3), defined as the median from the following three values

$$D(i, j - 1), D(i - 1, j), D(i, j - 1) + D(i - 1, j) - D(i - 1, j - 1) \tag{38}$$

and the FBH(3,7) predictor, which is the MAE-optimal FIR-Boolean hybrid filter, composed of a Boolean function with 7 inputs, the inputs being the outputs of the seven JPEG standardized predictors (which have at most 3 pixels as input).

From Table 2, the lowest average entropy for the prediction errors is obtained for the last two columns, where FIR-Boolean hybrid filters are used for prediction. The optimal FIR-Boolean hybrid filter performs the best, and the median adaptive predictor and the EE-optimal Boolean predictor compete for the second and third places, respectively. Another important fact, which is not presented in the table, is that the optimal stack filter was obtained to be only one of the pixels in the prediction mask, independently of the other two. This may be explained by the high correlation which exists between the neighboring pixels in an image, and also by the fact that only 18 different nontrivial stack filters may be designed for 3-pixel windows. Note that the optimal Boolean predictor ranks first in three cases (for images Boats, Gold and Hotel) performing equally well as FIR-Boolean hybrid predictors.

Table 2. Experiment # 1. Entropy of the residual image for 3-pixels prediction mask, different predictors.

Image	No prediction	JPEG	Boolean(3)	Stack(3)	MedAP	FBH(3,7)
Balloon	7.346	3.172	3.122	3.331	3.122	3.104
Barb1	7.555	5.302	5.208	5.343	5.208	5.203
Barb2	7.484	5.236	5.162	5.190	5.182	5.124
Boats	7.088	4.469	4.313	4.673	4.313	4.313
Girl	7.288	4.225	4.318	4.496	4.210	4.180
Gold	7.530	4.875	4.717	5.089	4.717	4.717
Hotel	7.546	4.943	4.735	5.122	4.735	4.735
Zelda	7.334	4.179	4.106	4.161	4.116	4.062
<b>Average</b>	<b>7.396</b>	<b>4.550</b>	<b>4.460</b>	<b>4.676</b>	<b>4.450</b>	<b>4.430</b>

As a conclusion of this experiment, it can be stated that optimally designed Boolean and FIR-Boolean hybrid predictors are promising alternatives to linear prediction, their properties being further analyzed in next experiments.

### Experiment #2. Using different prediction masks for Boolean and Stack predictors.

This experiment analyzes the efficiency of EE-optimally designed Boolean predictors and that of MAE-optimal stack predictors, for nesting prediction masks with sizes from 3 to 10 pixels.

The entropy of residual images for various prediction masks is presented in Table 3 for EE-optimal Boolean predictors and in Table 4 for MAE-optimal stack predictors. Many different shapes for the prediction mask may be examined, but here the selection is made as to always enlarge an existing prediction mask, by adding one more pixel. The added pixel is always one of the nearest (in Euclidean metrics) to the predicted one. It can be noted for the case of EE-optimal Boolean predictors that the average entropy of the residual images decreases with 3.5% when the size of the prediction mask increases from 3 to 10 pixels. However, there are images (e.g. Girl) for which this decrease is more important (about 8%), and other images (e.g. Gold) for which the entropy of the error is almost constant.

For the case of optimal stack predictors, enlarging the prediction mask to more than 5 pixels produces no further improvement in reducing the entropy of the prediction errors.

For each prediction mask, EE-optimal Boolean predictor results in a smaller entropy for the residual image than that obtained with the MAE-optimal stack predictor.

From this experiment it may be concluded that enlarging the prediction mask has the potential to reduce the entropy of the residual image for EE-optimal Boolean predictors.

### Experiment #3. Optimal FIR-Boolean hybrid predictors with different numbers of linear combinations in the first stage.

This experiment analyzes the performances of MAE-optimal FIR-Boolean hybrid predictors when the linear predictors in the first stage are fixed, and the optimization is performed for the Boolean filter in the second stage. Different numbers of linear predictors are considered in the first stage, selected from the JPEG standard predictors.

Table 3. Experiment # 2. Entropy of the residual image obtained using EE-optimal Boolean predictors with different prediction masks.

Image								
Balloon	3.122	3.122	3.122	3.122	3.122	3.055	3.051	3.018
Barb1	5.208	5.208	5.189	5.143	5.106	5.050	5.013	4.973
Barb2	5.162	5.162	5.140	5.126	5.079	5.069	5.051	5.036
Boats	4.313	4.313	4.313	4.313	4.313	4.296	4.272	4.265
Girl	4.318	4.318	4.318	4.318	4.254	4.112	4.070	4.055
Gold	4.717	4.717	4.717	4.717	4.717	4.705	4.696	4.696
Hotel	4.735	4.735	4.730	4.697	4.679	4.638	4.625	4.625
Zelda	4.106	4.089	4.035	4.013	3.973	3.947	3.923	3.918
<b>Average</b>	<b>4.460</b>	<b>4.458</b>	<b>4.445</b>	<b>4.431</b>	<b>4.405</b>	<b>4.359</b>	<b>4.338</b>	<b>4.323</b>

Table 4. Experiment # 2. Entropy of the residual images for MAE-optimal stack predictors with different prediction masks. The average figures of merit are computed for the set of eight test images in Experiment # 1.

Image								
Balloon	3.331	3.225	3.225	3.225	3.225	3.225	3.225	3.225
Barb2	5.190	5.190	5.190	5.190	5.190	5.190	5.183	5.183
Girl	4.496	4.437	4.435	4.435	4.435	4.435	4.435	4.435
Hotel	5.122	5.065	4.985	4.985	4.984	4.986	4.985	4.985
<b>Average</b>	<b>4.676</b>	<b>4.624</b>	<b>4.602</b>	<b>4.604</b>	<b>4.604</b>	<b>4.604</b>	<b>4.602</b>	<b>4.602</b>

Table 5. Experiment #3. Entropy of the residual images for MAE-optimal FBH predictors, for a variable number of linear combinations. The average figures are computed for the set of eight test images in Experiment # 1.

Image	FBH(3, 7)	FBH(3, 6)	FBH(3, 5)	FBH(3, 4)	FBH(3, 3)
Balloon	3.104	3.103	3.100	3.105	3.122
Barb2	5.124	5.124	5.147	5.147	5.182
Girl	4.180	4.180	4.187	4.205	4.210
Hotel	4.735	4.735	4.735	4.735	4.735
<b>Average</b>	<b>4.430</b>	<b>4.430</b>	<b>4.434</b>	<b>4.437</b>	<b>4.450</b>

The entropy of the prediction errors for different prediction structures is presented in Table 5. The linear stage of the FIR-Boolean hybrid filters is as follows:

$$\begin{aligned}
\text{FBH}(3,3) &- \text{FIR}_1, \text{FIR}_2, \text{FIR}_4 \\
\text{FBH}(3,4) &- \text{FIR}_1, \text{FIR}_2, \text{FIR}_4, \text{FIR}_7 \\
\text{FBH}(3,5) &- \text{FIR}_1, \text{FIR}_2, \text{FIR}_4, \text{FIR}_5, \text{FIR}_7 \\
\text{FBH}(3,6) &- \text{FIR}_1, \text{FIR}_2, \text{FIR}_4, \text{FIR}_5, \text{FIR}_6, \text{FIR}_7 \\
\text{FBH}(3,7) &- \text{FIR}_1, \text{FIR}_2, \text{FIR}_3, \text{FIR}_4, \text{FIR}_5, \text{FIR}_6, \text{FIR}_7
\end{aligned}$$

where  $\text{FIR}_1, \dots, \text{FIR}_7$  denote the seven JPEG predictors:

$$\begin{aligned}
Y_{\text{FIR}_1}(i, j) &= D(i, j - 1) \\
Y_{\text{FIR}_2}(i, j) &= D(i - 1, j) \\
Y_{\text{FIR}_3}(i, j) &= D(i - 1, j - 1) \\
Y_{\text{FIR}_4}(i, j) &= D(i, j - 1) + D(i - 1, j) - D(i - 1, j - 1) \\
Y_{\text{FIR}_5}(i, j) &= D(i, j - 1) + (D(i - 1, j) - D(i - 1, j - 1))/2 \\
Y_{\text{FIR}_6}(i, j) &= D(i - 1, j) + (D(i, j - 1) - D(i - 1, j - 1))/2 \\
Y_{\text{FIR}_7}(i, j) &= (D(i - 1, j) + D(i, j - 1))/2
\end{aligned}$$

From Table 5 it may be observed that adding more linear combinations to the first filtering stage will reduce the prediction error entropy. However, this reduction is less important than that obtained in the previous experiment for Boolean filters with increasing prediction mask sizes.

It was observed that the linear combinations  $\text{FIR}_4$  and  $\text{FIR}_7$  achieve the best prediction performances when included as linear substructures in an FBH filtering architecture for prediction.




An important conclusion is reached at this point: optimal FIR-Boolean Hybrid filters achieve the best performance among the proposed predictors for 3-pixel masks.

**Experiment #4. Block-optimal Boolean predictors for various numbers of blocks.** The performance of block-optimal Boolean predictors is investigated in this experiment. The image is divided into small rectangular regions and an MAE-optimal Boolean predictor is designed for each region.

The average values of the cumulative entropy criterion over all eight images are presented in Table 6, for different prediction masks and different block sizes. A significant reduction in the error entropy is observed in Table 6 only for block sizes smaller than  $64 \times 64$  pixels. As the block size is reduced the overhead necessary to encode a large number of filters increases. The tradeoff between these two aspects is found at the minimum of the global criterion  $\mathcal{R}$ . For small prediction masks (e.g. 3 pixels) the lowest average global criterion is obtained for  $16 \times 16$  block sizes. When the prediction mask increases, the minimum value of the global criterion is obtained for larger blocks (e.g.  $32 \times 32$  block-size for 6-pixel mask, or even the whole image for 10-pixel mask).

From Table 6 it can be concluded that block-optimal Boolean prediction with equally-sized blocks is a good solution for small prediction masks. Block-optimal Boolean prediction will

Table 6. Experiments #4 and #5. Average entropy of the residual images ( $H$ ), bitrates for predictors encoding ( $BP=2^N \cdot n_b/T$ , see Equation (36)) and global criterion  $\mathcal{R}$  for MAE block-optimal Boolean and FIR-Boolean hybrid prediction for different prediction masks (the results are averages over the eight test images).

Size of block		Boolean predictors			FIR-Boolean hybrid predictors		
					FBH(3,3)	FBH(3,4)	FBH(6,8)
$576 \times 720$	$H$	4.518	4.482	4.359	4.450	4.437	4.346
	BP	2.e-5	1.5e-4	0.002	2.e-5	4.e-5	0.001
	$\mathcal{R}$	<b>4.518</b>	<b>4.482</b>	<b>4.361</b>	<b>4.450</b>	<b>4.437</b>	<b>4.347</b>
$128 \times 128$	$H$	4.519	4.422	4.295	4.423	4.392	4.268
	BP	6.e-4	0.004	0.074	6.e-4	0.001	0.018
	$\mathcal{R}$	<b>4.519</b>	<b>4.426</b>	<b>4.369</b>	<b>4.423</b>	<b>4.393</b>	<b>4.284</b>
$64 \times 64$	$H$	4.491	4.378	4.223	4.410	4.367	4.222
	BP	0.002	0.017	0.267	0.002	0.004	0.067
	$\mathcal{R}$	<b>4.493</b>	<b>4.396</b>	<b>4.490</b>	<b>4.412</b>	<b>4.371</b>	<b>4.289</b>
$32 \times 32$	$H$	4.452	4.314	4.089	4.387	4.330	4.154
	BP	0.008	0.064	1.022	0.008	0.016	0.255
	$\mathcal{R}$	<b>4.460</b>	<b>4.378</b>	<b>5.111</b>	<b>4.395</b>	<b>4.346</b>	<b>4.409</b>
$16 \times 16$	$H$	4.404	4.216		4.356	4.286	4.027
	BP	0.031	0.250		0.031	0.062	1.
	$\mathcal{R}$	<b>4.435</b>	<b>4.466</b>		<b>4.387</b>	<b>4.348</b>	<b>5.027</b>
$8 \times 8$	$H$	4.340	4.039		4.303	4.220	
	BP	0.125	1.		0.125	0.250	
	$\mathcal{R}$	<b>4.465</b>	<b>5.039</b>		<b>4.428</b>	<b>4.470</b>	
$4 \times 4$	$H$	4.217			4.192	4.082	
	BP	0.5			0.5	1.	
	$\mathcal{R}$	<b>4.717</b>			<b>4.692</b>	<b>5.082</b>	

not improve the coding performance for larger prediction masks. Moreover, using many block-optimal Boolean filters with small prediction masks (on many blocks) will result in values for the performance criterion  $\mathcal{R}$  close to those obtained when only one optimal Boolean filter with a large prediction mask (for the whole image) is used.

**Experiment #5. Block-optimal FBH predictors for various numbers of blocks.** The performance of MAE block-optimal FBH filters, with fixed linear substructures, is investigated in this experiment. Table 6 presents the block-optimal FBH prediction performances which are the counterpart of (one-block)-optimal FBH prediction performed in Experiment #3. Additionally, a new FIR-Boolean hybrid structure, FBH(6,8), is considered by enlarging the 6 pixel prediction mask (see Table 3) of the Boolean filter with two more entries, the outputs of linear predictors  $FIR_4$  and  $FIR_7$  from JPEG.

It can be observed that for small  $Q$ , the FBH( $N$ ,  $Q$ ) predictors have the minimum cumulative criterion for block sizes  $16 \times 16$ , or  $32 \times 32$ .

Table 7. Experiment #6. Cumulative criterion for different prediction masks for quadtree coding scheme, for adaptive block sizes between  $32 \times 32$  and  $4 \times 4$ .

Image	$\begin{smallmatrix} \circ & \circ \\ \circ & \bullet \end{smallmatrix}$	<b>Boolean</b> $\begin{smallmatrix} \circ & \circ & \circ \\ \circ & \bullet & \circ \end{smallmatrix}$	$\begin{smallmatrix} \circ & \circ & \circ & \circ \\ \circ & \circ & \bullet & \circ \end{smallmatrix}$	<b>FIR-Boolean hybrid</b>		
				FBH(3,3)	FBH(3,4)	FBH(3,5)
Balloon	3.120	3.061	3.070	3.089	3.017	3.024
Barb1	5.068	4.974	4.943	5.083	5.028	5.039
Barb2	5.081	5.044	5.046	5.071	5.051	5.065
Boats	4.306	4.275	4.275	4.236	4.214	4.214
Girl	4.218	4.152	4.163	4.164	4.112	4.108
Gold	4.744	4.725	4.715	4.699	4.651	4.662
Hotel	4.689	4.673	4.670	4.686	4.634	4.646
Zelda	4.085	3.945	3.942	3.992	3.933	3.947
<b>Average <math>\mathcal{R}</math></b>	<b>4.404</b>	<b>4.356</b>	<b>4.353</b>	<b>4.377</b>	<b>4.330</b>	<b>4.338</b>

When  $N$  increases, the minimum global criterion is obtained for larger blocks, and the block-optimal design procedure does not improve too much compared to (one-block)-optimal design procedure. However, it must be noted that the optimal values of the cumulative criterion are always smaller for FIR-Boolean hybrid filters than those for Boolean filters with the same prediction mask.

**Experiment #6. Adaptive size block-optimal prediction.** This experiment illustrates the performance of adaptive size MAE block-optimal Boolean and FIR-Boolean hybrid predictors.

Prediction masks with 3–5 pixels are used in this experiment since block-optimal prediction proved useful only for small prediction masks.

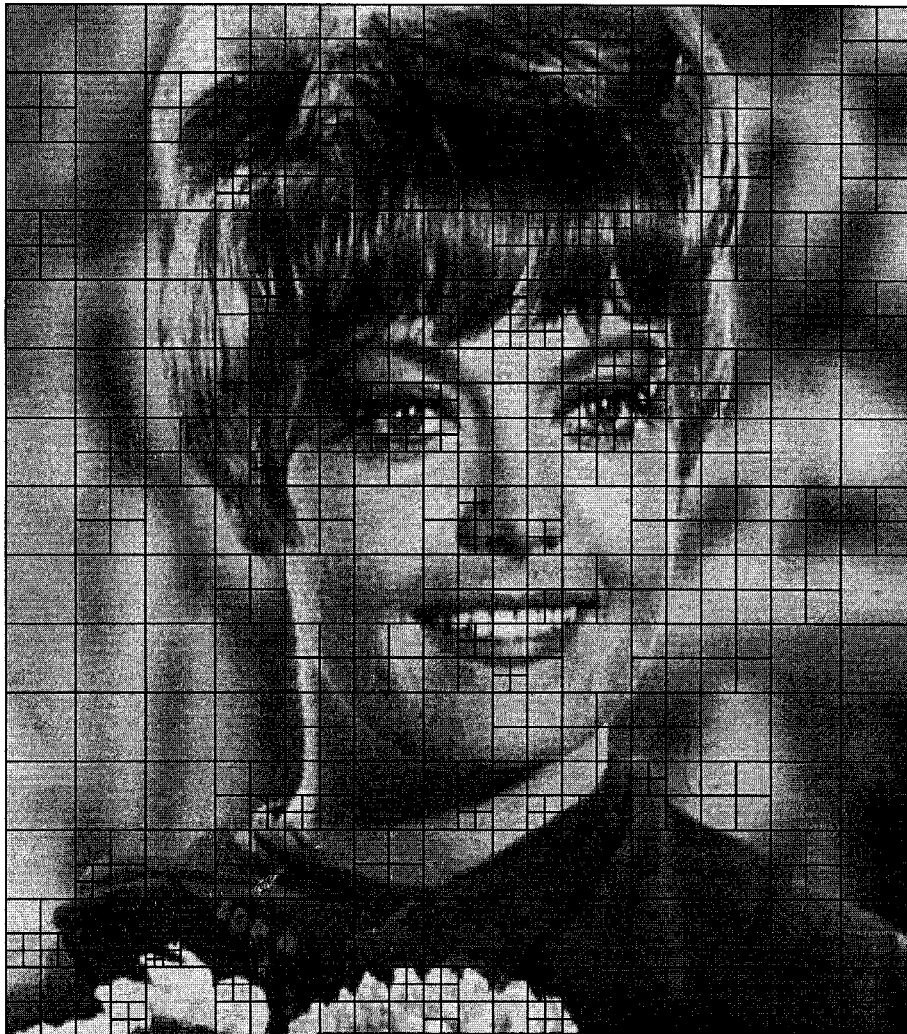
First, the image was split into regular blocks of size  $L \times L = 32 \times 32$ . Adaptive size splitting of these blocks is then performed using the procedure described in Section 3.2. The deepest level where splitting is allowed is set to  $4 \times 4$  pixel blocks.

The quadtree block dividing procedure is performed for all the  $L \times L$  blocks, but since the search for adaptive splitting is only sub-optimal, the procedure is repeated several times, starting from the partition found after a complete pass through the image. The iterations are stopped when no additional improvement in the cumulative criterion  $\mathcal{R}$  is obtained (2 to 4 iterations are usually required). The results presented in Table 7 show that the average values of the cumulative criterion are smaller than those when fixed block sizes are used, showing the usefulness of the proposed procedure.

### 5.3. Multiresolution Prediction

The performance of the multiresolution prediction scheme described in Section 4 is presented in this subsection. In all 3 stages, EE-optimal Boolean predictors and MAE-optimal FBH predictors are used.

For the case of optimal Boolean filters, the prediction/interpolation masks for each of the 3 stages are those presented in Figure 5. For the case of FIR-Boolean hybrid filters, 2



*Figure 6.* Final partition of image Zelda resulted using the adaptive size block-optimal Boolean prediction with 4-pixels prediction mask.

*Table 8.* Entropy of the prediction errors for multiresolution prediction schemes, for optimal Boolean and FIR-Boolean hybrid predictors and for a hierarchical interpolation scheme using linear filters.

<b>Image</b>	<b>Boolean(6)</b>	<b>FBH(6,8)</b>	<b>HINT</b>
Balloon	3.127	2.986	3.29
Barb1	5.211	4.934	5.41
Barb2	5.222	4.991	5.47
Boats	4.368	4.221	4.59
Girl	4.224	4.015	4.43
Gold	4.767	4.678	4.90
Hotel	4.755	4.639	5.04
Zelda	3.987	3.844	4.13
<b>Average</b>	<b>4.457</b>	<b>4.288</b>	<b>4.61</b>

linear combinations are added to the linear stage: in stage 1, linear predictors  $FIR_4$  and  $FIR_7$  from JPEG scheme are added to the 6 values in the prediction mask in Figure 5, and in stages 2 and 3 the averages of the 2 pairs of opposite pixels nearest to the predicted (interpolated) one are included. For comparison, the table contains also the entropy of the errors from HINT procedure presented in [5]. For all images, the optimal predictors introduced in this paper perform better than HINT scheme. When comparing the multiresolution prediction error entropy with that achieved by sequential prediction, the sequential prediction always surpasses the multiresolution prediction. The main source of inefficiency in the multiresolution prediction is determined by the high prediction errors in the first stage, which is the cost to be paid for a speedy reception of the low resolution image at the receiver.

## 6. Conclusions

The prediction performances of optimal Boolean, stack and FIR-Boolean hybrid predictors for lossless image compression have been investigated in this paper. New design procedures for Boolean filters have been introduced, where the optimality criterion is the Error Entropy (EE). The connections between the EE criterion and the MAE criterion were revealed for the case when the histogram of prediction errors is modeled using piecewise symmetrical exponentials.

We proposed new optimal algorithmic structures, whose parameters are fitted as much as possible to the image to be transmitted, and we showed that the cost of transmitting the predictor parameters is low enough to obtain an overall competitive lossless coding scheme. The use of the EE-optimal and MAE-optimal Boolean and stack filters in the sequential prediction structure was considered, under different instances: global-optimal, block-optimal, adaptive-size-block-optimal and multiresolution.

We experimented with images from the (standard) JPEG test image set to obtain a mean-

Table 9. Summary of the best performances: the entropy of prediction error for different predictors.

Image	JPEG	GAP	Boolean(10)		FBH(6,8)	FBH(6,8)
			MAE-optimal	EE-optimal	Sequential	Multiresolution
Balloon	3.172	3.046	3.100	3.018	3.024	2.986
Barb1	5.302	5.137	5.007	4.973	5.061	4.934
Barb2	5.236	5.190	5.056	5.036	5.047	4.991
Boats	4.469	4.286	4.293	4.265	4.256	4.221
Girl	4.225	4.103	4.108	4.055	4.075	4.015
Gold	4.875	4.676	4.708	4.696	4.690	4.678
Hotel	4.943	4.661	4.637	4.625	4.649	4.639
Zelda	4.179	3.951	3.964	3.918	3.968	3.844
<b>Average</b>	<b>4.550</b>	<b>4.363</b>	<b>4.359</b>	<b>4.323</b>	<b>4.346</b>	<b>4.288</b>

ingful comparison of our newly introduced procedure with other reported prediction methods.

We summarize in Table 9 the best prediction results obtained in our global prediction approach and compare them with the performances of standard algorithms (the best global and block JPEG [5]) and with the best prediction results we found in the literature (the recently proposed gradient adjusted predictor, used by CALIC system [36]).

The most important conclusions reached after our extensive experiments are the following:

- The best performance is obtained with block-optimal FIR-Boolean hybrid filters for small prediction masks of 3 to 6 pixels (the adaptive-size block-optimal prediction slightly outperforming regular block-optimal prediction). The FBH(6,8) predictor yields the average cumulative criterion  $\mathcal{R} = 4.284$  (see Table 6), ranking the first among all our predictors and any other predictor we found in the literature.
- Multiresolution methods have similar performance with sequential prediction methods and are useful since they provide progressive decoding methods, very convenient in interactive image transmission, e.g. in a multimedia environment.
- The best one-block sequential prediction performance is obtained by EE-optimal Boolean predictors with 10 pixels prediction mask.

These results encourage us to further investigate the lossless image coding applications of optimal Boolean and FIR-Boolean hybrid filters, namely how the optimal prediction stage must be combined with the most suitable error coding method.

Since the recent research in the area of lossless image coding focuses on new classes of prediction error *coding* methods, we currently investigate the combination of our optimal prediction stage with context error coding methods, in order to obtain very efficient lossless coding schemes.

## References

1. T. Endoh and Y. Yamazaki, "Progressive Coding Scheme for Multilevel Images," In *Proc. Picture Coding Symposium*, 1986, pp. 21–22.
2. A. K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
3. M. Rabbani and P. W. Jones, *Digital Image Compression Techniques*, Bellingham, Washington: SPIE Opt. Eng. Press, 1991.
4. R. B. Arps and T. K. Truong, "Comparison of International Standards for Lossless Still Image Compression," *Proceedings of the IEEE*, vol. 82, no. 6, 1994, pp. 889–899.
5. N. D. Memon and K. Sayood, "Lossless Image Compression: A Comparative Study," In *Proc. SPIE Still Image Compression*, vol. 2418, San Jose, California, 1995, pp. 8–20.
6. T. V. Ramabadran and K. Chen, "The Use of Contextual Information in the Reversible Compression of Medical Images," *IEEE Transactions on Medical Imaging*, vol. 11, no. 2, 1992, pp. 185–195.
7. P. Roos, M. A. Viergever, M. C. A. Van Dijke, and J. H. Peters, "Reversible Intraframe Compression of Medical Images," *IEEE Transactions on Medical Imaging*, vol. 7, no. 4, Dec. 1988, pp. 328–336.
8. G. K. Wallace, "The JPEG Still Picture Compression Standard," *Communications of the ACM*, vol. 34, no. 4, 1991, pp. 31–44.
9. D. H. Kang, J. H. Choi, Y. H. Lee, and C. Lee, "Applications of a DPCM System with Median Predictors for Image Coding," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 3, Aug. 1991, pp. 429–435.
10. C.-C. Lien, C.-L. Huang, and I.-C. Chang, "A New ADPCM Image Coder Using Frequency Weighted Directional Filter," In *Proc. SPIE Visual Communications and Image Processing*, vol. 2501, Taipei, Taiwan, 1995, pp. 647–657.
11. St. Martucci, "Reversible Compression of HDTV Images Using Median Adaptive Prediction and Arithmetic Coding," In *Proc. IEEE International Symposium on Circuits and Systems ISCAS-90*, New Orleans, Louisiana, 1990, pp. 1310–1313.
12. N. D. Memon, V. Sippy, and X. Wu, "An Asymmetric Lossless Image Compression Technique," In *Proc. IEEE International Conference on Image Processing ICIP-95*, vol. III, Washington, D.C., 1995, pp. 97–100.
13. A. Said and W. A. Pearlman, "An Image Multiresolution Representation for Lossless and Lossy Compression," *IEEE Transactions on Image Processing*, vol. 5, no. 9, Sep. 1996, pp. 1303–1311.
14. P. D. Wendt, E. J. Coyle, and N. C. Gallagher, Jr., "Stack Filters," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 8, Aug. 1986, pp. 898–911.
15. J.-H. Lin and Y.-T. Kim, "Fast Algorithms for Training Stack Filters," *IEEE Transactions on Signal Processing*, vol. 42, April 1994, pp. 772–781.
16. J.-H. Lin, T. M. Sellke, and E. J. Coyle, "Adaptive Stack Filtering Under the Mean Absolute Error Criterion," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 6, June 1990, pp. 938–954.
17. I. Täbuş, D. Petrescu, and M. Gabbouj, "A Training Framework for Stack and Boolean Filtering—Fast Optimal Design Procedures and Robustness Case Study," *IEEE Transactions on Image Processing Special Issue on Nonlinear Image Processing*, vol. 5, no. 6, June 1996, pp. 809–826.
18. E. J. Coyle and J.-H. Lin, "Stack Filters and the Mean Absolute Error Criterion," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 8, Aug. 1988, pp. 1244–1254.
19. I. Täbuş, *Training and Model Based Approaches for Optimal Stack and Boolean Filtering with Applications in Image Processing*, Ph.D. dissertation, Tampere University of Technology, 1995.
20. E. J. Coyle, J.-H. Lin, and M. Gabbouj, "Optimal Stack Filtering and the Estimation and Structural Approaches to Image Processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 12, Dec. 1989, pp. 2037–2066.
21. M. Gabbouj and E. J. Coyle, "Minimum Mean Absolute Error Stack Filtering with Structural Constraints and Goals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 6, June 1990, pp. 955–968.

22. D. Petrescu, I. Täbuş, and M. Gabbouj, "Edge Detection Based on Optimal Stack Filtering under Given Noise Distribution," In *Proc. European Conference on Circuit Theory and Design ECCTD-95*, Istanbul, Turkey, Aug. 1995, pp. 1023–1026.
23. D. Petrescu, I. Täbuş, and M. Gabbouj, "Adaptive Skeletonization using Multistage Boolean and Stack Filtering," In *Proc. VII European Signal Processing Conference, EUSIPCO-94*, Edinburgh, United Kingdom, Sept. 1994, pp. 951–954.
24. B. Zeng, H. Zhou, and Y. Neuvo, "FIR Stack Hybrid Filters," *Optical Engineering*, vol. 30, no. 7, y 1991, pp. 965–975.
25. K. D. Lee and Y. H. Lee, "Threshold Boolean Filters," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, Aug. 1994, pp. 2022–2036.
26. M. Gabbouj, E. J. Coyle, and N. C. Gallagher, Jr., "An Overview of Median and Stack Filtering," *Circuits, Systems and Signal Processing*, vol. 11, no. 1, 1992, pp. 7–45.
27. M. Gabbouj, P. T. Yu, and E. J. Coyle, "Convergence Behavior and Root Signal Sets of Stack Filters," *Circuits, Systems and Signal Processing*, vol. 11, no. 1, 1992, pp. 171–194.
28. D. Petrescu, I. Täbuş, and M. Gabbouj, "FIR-Boolean Hybrid Filtering Architecture and Applications to Image Processing," In *Proc. Finnish Signal Processing Symposium FINSIG'97*, Pori, Finland, May 1997, pp. 124–128.
29. W. K. Pratt, *Digital Image Processing*, New York: Wiley, 1990.
30. I. Täbuş, D. Petrescu, and M. Gabbouj, "Training Based Optimal Stack Filter Design Under Structural Constraints," In *Proc. International Conference on Image Processing, ICIP-96*, Lausanne, Switzerland, Sept. 1996, pp. 761–764.
31. D. Petrescu, I. Täbuş, and M. Gabbouj, "Locally Adaptive Techniques for Stack Filtering," In *Proc. VIII European Signal Processing Conference, EUSIPCO-96*, Trieste, Italy, Sept. 1996, pp. 587–590.
32. J. Vaisey and A. Gersho, "Image Compression with Variable Block Size Segmentation," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, Aug. 1992, pp. 2040–2060.
33. H. Lee, Y. Kim, and S. Oh, "Lossless Compression of Medical Images by Prediction and Classification," *Optical Engineering*, vol. 33, no. 1, Jan. 1994, pp. 160–166.
34. A. Lehtonen and M. Renfors, "Nonlinear Quincunx Interpolation Filtering," In *Proc. SPIE Visual Communications and Image Processing*, vol. 1360, Lausanne, 1990, pp. 131–141.
35. B. Zeng and A. N. Venetsanopoulos, "A Comparative Study of Several Nonlinear Image Interpolation Schemes," In *Proc. SPIE Visual Communications and Image Processing*, vol. 1818, Boston, 1992, pp. 21–29.
36. X. Wu and N. D. Memon, "CALIC—A Context Based Adaptive Lossless Image Codec," In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-96*, vol. IV, Atlanta, U.S.A., May 1996, pp. 1890–1893.