

A ZERO-TREE LIKE CODEC USING NONLINEAR SIGNAL DECOMPOSITIONS ¹

J. L. Paredes

G. R. Arce

N. C. Gallagher

Department of Electrical Engineering
University of Delaware
Newark, DE 19716

ABSTRACT

In this paper a nonlinear signal decomposition for image compression is presented. It uses a pyramid multiresolution scheme which is similar to that found using wavelet sub-band decompositions. The self-similarity across the different scales of the nonlinear signal decomposition is then exploited by the SPIHT algorithm which is modified to match the new signal decomposition. The nonlinear decomposition produces better edge preserving and less blocking artifacts in the reconstructed image than traditional wavelet decomposition, specially at very low bit rates.

INTRODUCTION

Recently, numerous image codecs which exploit the self-similarity of the wavelet transform among the different scales have been introduced. The embedded zero-tree codec introduced by Shapiro in 1993 [1] and the Set Partitioning in Hierarchical Trees (SPIHT) [2] are examples of this class of codecs. Among all coding methods, these codecs have the best performance for low bit rate compression (i.e. 0.5 - 0.2 bpp). For very low bit rate compression (below 0.2 bpp), however, these coding methods produce severe artifacts leading to unacceptable performance.

In this paper we show that nonlinear signal decompositions are well suited for zero-tree like coding and that codecs utilizing nonlinear signal decomposition perform significantly better than traditional zero-tree like codecs using wavelet transforms, particularly at very low bit rates.

Our codec borrows its structure from the SPIHT algorithm but uses an order statistic signal decomposition instead of the traditional wavelet signal decomposition. Consequently, the algorithm must be modified such that different subbands must be weighted according to their influence on the reconstruction. The order in which the coefficients are encoded not only depends on their magnitude (as in the wavelet case) but also of their position within the different subbands.

The nonlinear filter bank used in the decomposition is based on the so called median affine filter [3] which can be tuned to the desired level of nonlinearity ranging from a

standard linear FIR filter to a median filter. We profit from the best attributes of both filtering methods in a tunable fashion. Thus the preservation of edges and details which are the best characteristics of the median filter as well as the effectiveness of the linear filter in smooth image regions can be exploited jointly.

NONLINEAR DECOMPOSITION

Let \mathbf{p} be the set of pixels that defines the original image such that $p_{i,j}$ represents the value of the element (pixel) of the original image in the coordinate (i, j) . We define the polyphase components of the original image as the subsets of pixels x_{00} , x_{01} , x_{10} , and x_{11} such that

$$\begin{aligned}x_{00} &= \{x_{00,i,j} = p_{i,j} \text{ for all } i\text{-even and } j\text{-even}\} \\x_{01} &= \{x_{01,i,j} = p_{i,j} \text{ for all } i\text{-even and } j\text{-odd}\} \\x_{10} &= \{x_{10,i,j} = p_{i,j} \text{ for all } i\text{-odd and } j\text{-even}\} \\x_{11} &= \{x_{11,i,j} = p_{i,j} \text{ for all } i\text{-odd and } j\text{-odd}\}\end{aligned}$$

As shown in Fig. 1, the nonlinear decomposition is obtained as follows. The original image, \mathbf{p} , is split into its polyphase components x_{00} , x_{01} , x_{10} , and x_{11} as above. Next, the “low-low” subband Y_{00} is simply x_{00} , a subsampled version of the original image. The x_{00} polyphase is used to predict the x_{11} polyphase, thus, the prediction error defines the “high-high” subband, i.e. $Y_{11} = x_{11} - \hat{x}_{11}$, where $\hat{x}_{11} = \mathcal{F}(x_{00})$ is the predicted value of x_{11} from x_{00} samples. Both components x_{00} and x_{11} are used to predict the x_{01} and x_{10} components and the difference between the real and predicted values form the “high-low” and “low-high” subbands respectively [4]. Fig. 2 shows graphically the nonlinear decomposition.

The structure shown in Fig. 1 is similar to that presented in [5, 6], hence, itself guarantees perfect reconstruction from the subband coefficients regardless of the filter used. Thus the original polyphase components as function of the subbands are given by:

$$\begin{aligned}x_{00} &= Y_{00} \\x_{11} &= Y_{11} + \mathcal{F}(x_{00}) \\x_{01} &= Y_{01} + \mathcal{F}(x_{00}, x_{11}) \\x_{10} &= Y_{10} + \mathcal{F}(x_{00}, x_{11})\end{aligned}$$

It is clear that the choice of the filter \mathcal{F} is very important. In our simulations we used the median affine filter [3] since it can be suitably adapted to the desired level of

¹ This research has been supported through collaborative participation in the Advanced Telecommunications/Information Distribution Research Program (ATIRP) Consortium sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0002.

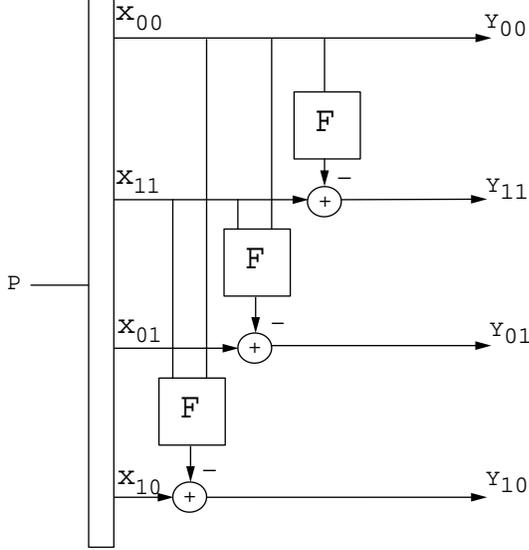


Figure 1: Nonlinear filter decomposition.

nonlinearity ranging from a standard linear FIR to a median filter. Thus, it is possible to select the most appropriate filter so that it yields the best energy compaction in the decomposition. The output of the median affine filter is given by:

$$\hat{x}_{11} = \frac{\sum_{i,j \in \mathcal{W}} g\left(\frac{x_{00,i,j} - x_{med}}{\gamma}\right) w_{i,j} x_{00,i,j}}{\sum_{i,j \in \mathcal{W}} g\left(\frac{x_{00,i,j} - x_{med}}{\gamma}\right) w_{i,j}}, \quad (1)$$

where \mathcal{W} is the observation window which contains the samples used to get the prediction of x_{11} , x_{med} is the sample median and $w_{i,j}$ are the prediction weights. The function $g(\cdot)$ is called the affine function and its main objective is to give a measure of the proximity of each sample in the observation window (\mathcal{W}) to the sample median. Thus, those samples that are close to the sample median are assigned a high affinity (≈ 1), whereas, those that are distant to the sample median are assigned a low affinity (≈ 0). The affine function is constrained to be unimodal and to have a maximum value of one at the origin. There are many functions that satisfy these requirements, one of them which was used in our simulations is the Gaussian affinity function $exp(-\frac{(x-\mu)^2}{\gamma})$, where γ is a tuning parameter which determines the desired level of linearity. Note that as $\gamma \rightarrow +\infty$ the median affine predictor behaves like a linear predictor and as $\gamma \rightarrow 0$ its behavior is like a median predictor. Therefore, in a tunable fashion, it is possible to exploit the best qualities of both filtering methods.

The recursion of the nonlinear filter bank over the “low-low” subband, Y_{00} , generates a pyramid like wavelet decomposition. In Fig. 3, we show a three level subband decomposition found using nonlinear filter bank and wavelet.

MODIFIED SPIHT

Once the decomposition is obtained, the resultant coefficients are passed through the SPIHT codec. This codec

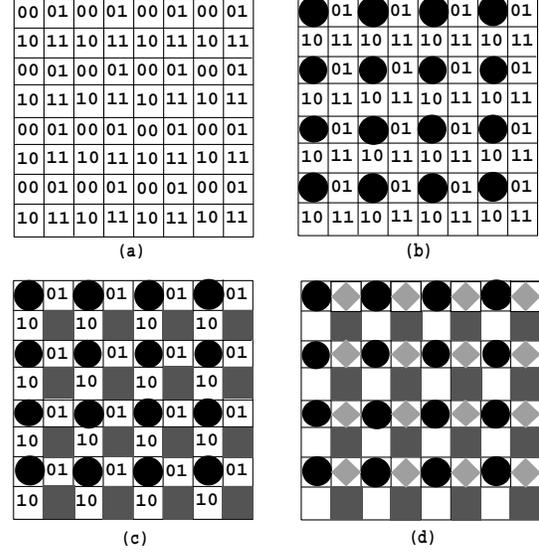


Figure 2: In Fig. 2a the polyphase components x_{00} , x_{01} , x_{10} , x_{11} are shown. In Fig. 2b the x_{00} polyphase (black dots) are used to predict the x_{11} polyphase (grey squares) of Fig. 2c. In Fig. 2d the x_{00} and x_{11} polyphases are used to predict the x_{01} (grey rhombus) and x_{10} (white squares) polyphases.

is designed to exploit: (1) Energy compaction, and (2) The self-similarity between subbands such that the coefficients are expected to be magnitude ordered if we move downward in the subband decomposition following the same spatial orientation [2].

This algorithm, however, is originally based on the wavelet decomposition, hence, the importance in the encoding of the coefficients is determined only by the magnitude of the coefficients. This follows as a result that the Euclidean norm is invariant to unitary transformations. Since the nonlinear transforms used here, are not unitary, the error norm in the transformed space will not be necessarily the same in the reconstructed space. Hence, the magnitude of the coefficients as well as their position in the pyramid decomposition must be taken into account in the coding stage. For this reason, the SPIHT codec must be modified according to the characteristics of the nonlinear decomposition such that those coefficients that yield the largest distortion reduction are given more priority in the encoding stage.

The structure of the proposed nonlinear decomposition is such that an error in a coefficient in the highest level of the pyramid will have more influence in the reconstructed image than an error of equal and possible larger amplitude in any other level. Moreover, at the same level, except in the highest level, the “high-high” subband which is used to predict the x_{01} and x_{10} polyphase components of the next level, introduces more distortion than that introduced by an error in the corresponding “high-low” and “low-high” subbands at the same decomposition level. Thus, those subbands that are more important in the reconstruction, as related to their influence in reducing the distortion of the reconstructed image are given more priority. As a measure



(a) Nonlinear filter decomposition.



(b) Wavelet decomposition.

Figure 3: Three level subband decompositions.

of distortion we use the peak signal-to-noise ratio PSNR defined as:

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \text{ dB}, \quad (2)$$

where MSE defines the mean square error between the original and reconstructed images given by:

$$MSE = \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M (p_{i,j} - \hat{p}_{i,j})^2, \quad (3)$$

with M and N as the size of the original image.

In order to minimize this error a certain rule of importance should be established in the encoding of each subband. One way to define this importance is to assign to each subband, a weight according to its influence in the reconstruction.

Error subband	$PE_{l,m}^{(n)}$	Weights
$Y_{00}^{(3)}$	21.44	1.00
$Y_{11}^{(3)}$	10.95	0.71
$Y_{01}^{(3)}$	5.72	0.52
$Y_{10}^{(3)}$	5.72	0.52
$Y_{11}^{(2)}$	3.12	0.38
$Y_{01}^{(2)}$	1.84	0.29
$Y_{10}^{(2)}$	1.84	0.29
$Y_{11}^{(1)}$	1.25	0.24
$Y_{01}^{(1)}$	1.00	0.22
$Y_{10}^{(1)}$	1.00	0.22

Table 1: Propagation error for a three level signal decomposition and weighted values for each subband.

To find such weights, let $\Delta Y_{l,m}^{(n)}$ be an error in a coefficient in the (l, m) subband at the n_{th} level of the decomposition, where $(l, m) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ representing the low-low, high-low, low-high and high-high subbands respectively, the decomposition level is indexed as $n = 1, 2, 3, \dots, n'$ with n' as the last level of the decomposition.

The error in a reconstructed pixel that is affected by $\Delta Y_{l,m}^{(n)}$ can be expressed as:

$$\Delta p_{i,j} = p_{i,j} - \hat{p}_{i,j} = k_{i,j} \Delta Y_{l,m}^{(n)} \quad (4)$$

where $p_{i,j}$ and $\hat{p}_{i,j}$ are the original and reconstructed pixels respectively, and $k_{i,j}$ is a parameter that measures the influence of the $\Delta Y_{l,m}^{(n)}$ on the pixel $p_{i,j}$, given by:

$$k_{i,j} = \frac{\partial p_{i,j}}{\partial Y_{l,m}^{(n)}}. \quad (5)$$

Of course, this parameter depends on the proximity of the pixel $p_{i,j}$ to the error coefficient $\Delta Y_{l,m}^{(n)}$. Hence, for those pixels distant from the location of $\Delta Y_{l,m}^{(n)}$, and thereof not affected, this parameter is equal to zero.

When there is only one error, $\Delta Y_{l,m}^{(n)}$, the MSE given by (3) can be expressed as a function of these parameters $k_{i,j}$ as:

$$MSE = \frac{(\Delta Y_{l,m}^{(n)})^2}{MN} \sum_{i=1}^N \sum_{j=1}^M k_{i,j}^2. \quad (6)$$

The term

$$PE_{l,m}^{(n)} = \sum_{i,j \in S} k_{i,j}^2 \quad (7)$$

is called the propagation error of the subband (l, m) at the n_{th} level of the decomposition, and S is the set of pixels that are affected by the error $\Delta Y_{l,m}^{(n)}$. This propagation error is identical for all the coefficients in a same subband and level, except possibly for the coefficients on the boundary between consecutive subbands.

Table 1 shows the propagation error found for a three level signal decomposition using a mean filter with 2×2 window size. From this table, one can see for example that



Figure 4: Original image. 256 x 256 pixels. 8 bits per pixel.

at level-3, the low-low subband is more important than the high-high, but the high-high is more important than the low-high and high-low.

As an example of how the $PE_{l,m}^{(n)}$ values were determined, consider a coefficient error in the high-high subband of the first decomposition level ($\Delta Y_{11}^{(1)}$). According to the decomposition, the $Y_{11}^{(1)}$ is used to determine the x_{01} and x_{10} components of the original image, hence, an error in any coefficient (i, j) in $Y_{11}^{(1)}$ will affect only the reconstruction of its four pixels closest to it. For any of these neighboring pixels we have:

$$p_{i,j} = Y_{10}^{(1)}(i, j) + 1/4[Y_{00}^{(1)}(i, j) + Y_{00}^{(1)}(i+1, j) + Y_{11}^{(1)}(i, j) + Y_{11}^{(1)}(i, j+1)]$$

$$k_{i,j} = \frac{\partial p_{i,j}}{\partial Y_{11}^{(1)}} = 1/4.$$

$$\text{Using } PE_{l,m}^{(n)} = \sum \sum_{i,j \in S} k_{i,j}^2,$$

$$PE_{11}^{(1)} = (1/4)^2 + (1/4)^2 + (1/4)^2 + (1/4)^2 + (1)^2 = 1.25$$

The term $(1)^2$ includes the error $Y_{11}^{(1)}$ itself. In the same way an error in either $Y_{01}^{(1)}$ or $Y_{10}^{(1)}$ does not affect any other pixels that itself, hence, $PE_{10}^{(1)} = PE_{01}^{(1)} = 1$.

In order to find the weights that must be assigned to each subband, let $\Delta Y_{00}^{(n')}$ be a coefficient error in the low-low subband and $\Delta Y_{l,m}^{(n)}$ be a coefficient error in any other subband. These two errors produce the same distortion (same MSE) if the magnitudes of the errors are related as follows:

$$\frac{\Delta Y_{00}^{(n')}}{\Delta Y_{l,m}^{(n)}} = \sqrt{\frac{PE_{l,m}^{(n)}}{PE_{00}^{(n')}}} \quad (8)$$

where $PE_{00}^{(n')} > PE_{l,m}^{(n)}$ for all l, m and n ($l, m \neq 0, 0$ and $n \neq n'$). Thus, a coefficient error of magnitude $\Delta Y_{l,m}^{(n)}$ produces

the same distortion than that produced by a coefficient error in the low-low subband of magnitude $\sqrt{\frac{PE_{l,m}^{(n)}}{PE_{00}^{(n')}}} \Delta Y_{l,m}^{(n)}$.



(a)



(b)

Figure 5: Compression rate = 0.1bpp. (a) Standard SPIHT. (b) Nonlinear SPIHT.

This suggests that if each subband (l, m) at the n_{th} level of the decomposition is weighted by:

$$W_{l,m}^{(n)} = \sqrt{\frac{PE_{l,m}^{(n)}}{PE_{00}^{(n')}}} \quad (9)$$

the MSE is minimized, where $PE_{l,m}^{(n)}$ is the propagation error of an error in the subband (l, m) at the n_{th} level of the decomposition and $PE_{00}^{(n')}$ is the propagation error of an error in the low-low subband. Table 1 shows the different weights found for a three level decomposition.

Thus, before the transformed coefficients are encoded, these are weighted according to (9) at the different levels, i.e. $c'_{i,j} = W_{l,m}^{(n)} c_{i,j}$ where $c_{i,j}$ are the transform coefficients in the subband (l, m) . Of course, the nonlinear signal decomposition must guaranty perfect reconstruction when no compression is done, hence, an inverse process is applied after the decoding operation. After decoding, the received coefficients are then weighted so as to emphasize the high frequency components. This same procedure was used with the standard SPIHT codec but the results were not favorable and thus it was not used in that case.

Although the above analysis was done when \mathcal{F} is constrained to be a linear filter, the weighting structure produces good results even when the filter introduces nonlinearities.



(a)



(b)

Figure 6: Compression rate = 0.15bps. (a) Standard SPIHT. (b) Nonlinear SPIHT.

SIMULATIONS

In order to evaluate the proposed approach, we compare the results obtained with the new algorithm (nonlinear SPIHT) with those obtained by the standard SPIHT algorithm. Fig. 4 shows the original test image which was compressed to 0.1 and 0.15 bit per pixel (bpp). The coded images at 0.1 bpp and .15 bpp are shown in Figs. 5 and 6 respectively. The tuning parameter, (γ) , of the median affine filter was fixed to 1 which yields in general, good performance.

As shown in Figs. 5 and 6 the ringing artifacts around edges with the standard SPIHT algorithm are very disturbing, particularly at compression rates 0.1 bpp. These artifacts are not present with the nonlinear SPIHT coded, moreover, the edges are better defined in the nonlinear SPIHT than in the standard SPIHT. However, the standard SPIHT yields a better performance in the PSNR measure sense than the nonlinear SPIHT does (the PSNR in the standard SPIHT is approximately 1.5 dB larger than that in the Nonlinear case). We believe that at very low bit rates the most appropriate measure of performance is subjective and, in general, MSE based measures are not good indicators of image quality.

CONCLUSIONS

We have presented a nonlinear signal decomposition which uses a recursive nonlinear filter banks to get a pyramid decomposition similar to that found using wavelet. The self-similarity across different scales is then exploited by the SPIHT algorithm which was suitably modified according to the nonlinear decomposition yielded better image compression results at low bit rates.

1. REFERENCES

- [1] J. M. Shapiro. "Embedded image coding using zerotrees of wavelet coefficients". *IEEE Trans. Signal Processing*, Vol. 41:3445–3462, Dec. 1993.
- [2] A. Said and W. A. Pearlman. "A new fast and efficient image codec based on set partitioning in hierarchical trees". *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 6, June 1996.
- [3] A. Flaig, G. Arce, and K. Barner. "Affine order statistic filters". *Proceedings of the 1997 ICASSP Conference, Munich, Germany*, April 1997.
- [4] D. Florencio and R. Shafer. "Perfect reconstructing nonlinear filter banks". *Proceedings of the 1994 ICIP Conference, Austin, TX*, Vol. 2, Nov. 1994.
- [5] F. J. Hampson and J.C. Pesquet. "A nonlinear sub-band decomposition with perfect reconstruction". *Proceedings of the 1996 ICASSP Conference, Atlanta, GA*, Vol. 3:1523–1526, May 1996.
- [6] G.R. Arce, J.L. Paredes, N.C. Gallagher, and D. Lau. "Nonlinear signal decompositions for scalable image compression". *Proceedings of the 1997 ATIRP Conference, College Park, MD*, Vol. 1:137–141, Jan. 1997.