HIERARCHICAL IMAGE CODING WITH PYRAMID DPCM

Ricardo L. de Queiroz

João B. T. Yabu-uti

Depto de Comunicações - Faculdade de Engenharia Elétrica
Universidade Estadual de Campinas
CP 6101 Campinas - SP 13081
Fax : +55-192-394717 Phone : +55-192-397750
BRAZIL

Abstract :

Pyramid Coding is based upon the representation of an original image into sub-images with different resolutions. Those resulting images are separately coded with reduction of required entropy for perfect reconstruction. We will apply DPCM to Pyramid Coding strategy, and show that for practical quantizers, lower bit rates than regular DPCM are required for high quality resulting images. Also the Pyramid DPCM is relatively simple, feasible and allows a processing interval between samples that is longer than the sampling interval. Furthermore we will comment a blockwise scheme for multiresolution, which could largely decrease the coding bit rate.

1) INTRODUCTION

The main goal in studying image coding techniques is the compression of the required amount of data for its representation. Here we have intended to modify standard DPCM coding strategies [1] in order to deal with the short sampling interval in video signals. This period commonly lies near 100ns, but in the proposed scheme, adding parallel computation, it will increase to 200ns and 400ns, maintaining quality and reducing bit-rate. It allows frequency differenciated coding, like Sub-Band Coding [1], multiresolution, progressive transmission as well as DPCM simplicity. Due to this fact, we do not expect great saving , but considerable performance improvement when compared with regular DPCM under same conditions. That led us to a comparative behavior throughout this paper.

Since we seek for simplicity, we will be limited to unidimensional signals, which are obtained by raster scannig of the image over interlaced fields, as in commercial TV. The results could be easily upgraded in order to better exploit correlation of bidimensional signals. The test images are monochromatic and have been grabbed with 10 M samples/s under 8-bit quantization.

2) PYRAMID DPCM

2.1) Pyramid Coding

The pyramid technique for progressive coding of images can be found in [2], [3] and [4]. We will briefly describe the essence of the method.

Let the original image be an array of NxN pels, with N a power of 2. Here, we shall use the unidimensional sequence x(n) (which represents the full image by scanning it line by line). Let $x_0(n)=x(n)$, prefilter this sequence with $\pi f_s/2$ and subsample it with half the original sampling rate f_s. Call this poorer image $x_1(n)$. From $x_1(n)$, find $x'_1(n)$ by upsampling and interpolation, this will be an approximation of x(n). The array of differences between them can be found by $L_0(n)=x_0(n)-x'_1(n)$. If the prefilters and interpolators were ideal, Lo(n) would be the difference between the original image and the lowpass version version of it. Repeat the process finding $x_2(n)$ and $x'_2(n)$ and so on. In a general formulation, we have

$$L_k(n) = x_k(n) - x'_{k+1}(n) \qquad k = 0,1,\dots,M-1 \qquad \text{(E1)}$$

as the difference between an image with $1/2^k$ of the original resolution and an approximation of it. If $N^2=2^M$, then $x_M(n)$ is a single pixel image roughly corresponding to the mean of the original one.

Recovering x(n) from $x_M(n), L_0(n), L_1(n), \dots, L_{M-1}(n)$ would be an easy task, by reverting the process. Therefore the general decoding formulation is

$$x_k(n) = x'_{k+1}(n) + L_k(n) \qquad k=M-1,M-2,\dots,1,0 \ \text{(E2)}$$

Calling $x_M(n)$ as $L_M(n)$, the pyramid is formed by the levels $L_K(n)$ displaced hierarchically, as in Figure 1. The total number of samples (nodes) recorded all over the pyramid amounts roughly to twice the number of pixels or

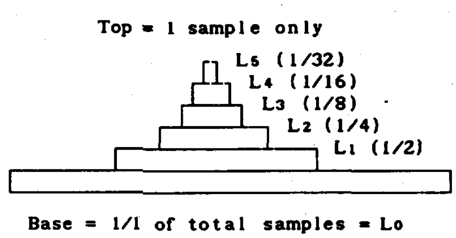


FIGURE 1 - Pyramid nodes levels

$$Nnodes = \sum_{i=0}^{M} N^2(2^{-i}) = 2(1-2^{-M})N^2. \quad (E3)$$

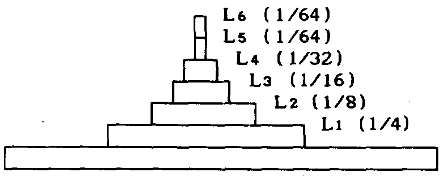
In order to reduce this number to N^2 the REDUCED-PYRAMID may be applied [4]. Another approach for reducing the pyramid would arise if no prefiltering is employed, therefore $x_i(n)$ would have half of its samples coincidents with those from $x_{i+1}(n)$ and $L_i(n)$ must record just those missing samples. Due to this approach, the levels would contain half the samples than non-reduced pyramid. Figure 2 sketches this new pyramid, with

$$Nnodes = N^2 2^{-M} + \sum_{i=1}^{M} N^2(2^{-i}) = N^2 \quad (E4)$$

This is an issue concerned about a trade-off between aliasing inaccuracy and elimination of overhead, but avoiding filtering processing and allowing any interpolator's length. This is a distinct technique than that found in [4] and just works for pyramids with few levels, into which aliasing would not be so determinant.

2.2) Applying DPCM to Reduced-Pyramids

In order to make the Pyramid feasible, we will restrict the number of levels to three (M=2) and L2(n) will contain one fourth of the original samples. L1 and Lo are composed by differences between lower level missing samples and interpolative predictions of them. If we code L2(n) with DPCM, its samples would be coded by differenciating the samples and extrapolative predictions of them [1]. The global coding process would form a hybrid inter/extrapolative scheme and we can say that Lo,L1 and L2 are coded via DPCM, being the latter level coded with extrapolative prediction and the formers with interpolative prediction.



Base = 1/2 of total samples = Lo

FIGURE 2 - Reduced-pyramid nodes levels

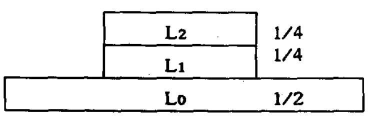


FIGURE 3 - 3-level Reduced-pyramid

In Figure 3 and 4 the steps towards the DPCM pyramid are illustrated, proceeding as follows :

i) Decimate, by 2 and 4, x(n) in order to find $x_1(n)$ e $x_2(n)$.

ii) Code $x_2(n)$ with a regular DPCM, therefore $L_2(n)$ is formed by the differences between $x_2(n)$ samples and their predictions.

iii) Interpolate $x_1(n)$ and let $L_0(n)$ be formed by the errors of this interpolation and x(n). Do not code the errors corresponding to samples represented in $x_1(n)$.

iv) Interpolate $x_2(n)$ and let $L_1(n)$ be formed by the errors of this interpolation and $x_1(n)$. Do not code the errors corresponding to samples represented in $x_2(n)$.

In Figure 5 a standard DPCM is represented with x(n) as input, e(n) as coded errors and $\tilde{x}(n)$ as locally decoded recontructed value of x(n). This coder is applied to the scheme of the Pyramid DPCM in Figure 6 . In this the filters are the interpolators and DEMUX2 and MUX2 are devices that divide and reconstruct, respectively, their input samples. (even-n samples for one branch and odd-n for the other).

Being α,β,γ the mean bit rates for L2's DPCM, L1 and Lo , the global bit rate produced by the Pyramid DPCM is given by :

$$R = (\alpha + \beta + 2\gamma)\frac{1}{4} \quad (E5)$$

3) BLOCKWISE PROGRESSIVE TRANSMISSION

Those three levels could be thought of as : L2 conveys information about a basic image, a blurred version of the original; L1 conveys information about a first refinement and Lo conveys the conclusive one, recomposing the original image.In many cases, there is no need for these refinements because images are composed by high and low frequency regions. Therefore we could divide the image into blocks, compute the SNR in each block and just refine regions where this ratio is below some threshold. In these blocks, we would not code Lo or even L1. We have prefiltered, decimated and interpolated the image "Zelda", obtaining images with 1/4 and 1/2 of the original bandwidth. These images were divided in 16x16 pels blocks and Figure 7 sketches the block density per SNR obtained. Note that there is a considerable amount of blocks whose SNR overpassed thresholds in the range 35-40 dB. Taking into account the relative frequency of over-threshold blocks in E5, we can rewrite it as

$$R = \frac{(\alpha + a\beta + 2b\gamma)}{4} + S \quad (E6)$$

where S is the side-information rate.

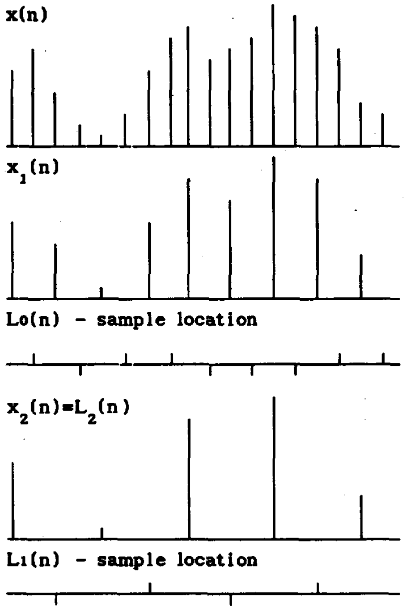


FIGURE 4 - Example.

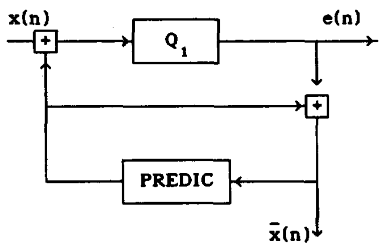


FIGURE 5 - DPCM Diagram

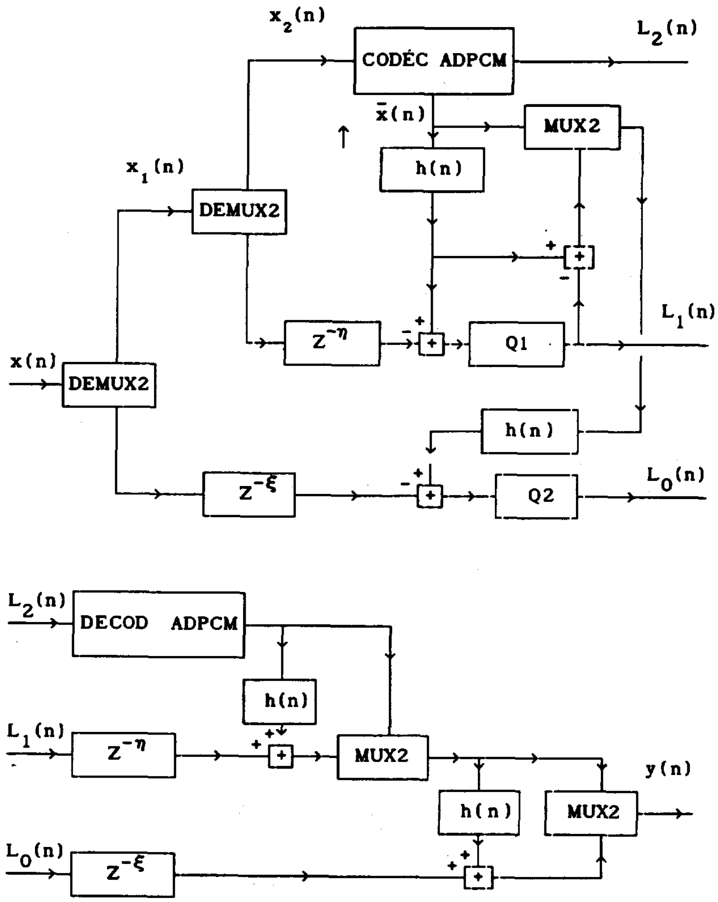


FIGURE 6 - Pyramid DPCM

4) REGULAR x PYRAMID DPCM

By comparing the standard and the Pyramid DPCM, we may say that the main advantages of the latter are

a)Better data compression, due to (i) the prediction gain of interpolation over extrapolation and (ii) to the fact that is possible to adopt three distinct quantization procedures, one for each level, optimizing coding and achieving improved subjective performance in comparison to the regular scheme.

b)Enlargement of processing interval between samples. Those intervals are 4 times longer for Lo and L1 and 2 times longer for Lo (See Figs. 4 and 6).

c)Facility to extend to a Multiresolution approach by conditional/progressive transmission

However, the main disadvantages rise from the needs for apropriated logic to multiplex those levels and from the addition of parallel computation (Fig. 6).

Now we must compare the predictors in both cases. The signal sent over the channel is composed by prediction errors of the samples x(n). Those predictions are achieved by discrete filtering applied to previous samples (extrapolative) or to advanced samples too (interpolative). Generally :

$\hat{x}(n)=[a_1 a_2 \ldots]\ [x(n-d_1)\ x(n-d_2)\ldots] = A^t X^t$ (E7)

Figure 8 depicts the interlaced-line-system. It is also shown the sample to be predicted and samples commonly used for intrafield prediction (X^t=[a b c] or X^t=[a]). Also, the predictors must be simple in order to not mask comparative results.

5) SIMULATIONS

For the simulations on still pictures, we used exactly the same DPCM for L_2 and for full-image. This includes adaptive 2D prediction and fixed 31 level scalar quantizer. For the prediction equations updated by a 2D LMS algorithm [9].

$$\mathbf{A}(n+1)=\mathbf{A}(n) + \mu\ (x(n)-\hat{x}(n))\ \mathbf{X}\ ;\ \mu = \frac{0.1}{255^2} \quad \text{(E14)}$$

The quantizer's laws for L_1 and L_0, here used were :

Range	Output	Range	Output	Range	Output
0-1	0	0-0	0	0-1	0
2-6	3	1-2	1	2-7	3
7-14	10	3-7	4	8-24	12
15-25	19	8-24	12	25-255	37
26-43	32	25-255	37		
44-255	55				

Figures 9-11 show comparative details extracted from reconstructed images ("Zelda", "Kitchen", "Girl"). In these, the upper left quarter (UL) is extracted from original image; upper right (UR): (L_0-7 levels); Bottom Left (BL) : (L_0-9 levels); Bottom Right (BR): Standard DPCM. L_1 was quantized into 11 levels. Those results are summarized in TABLE III, indicating overall SNR and global mean bit-rate. In the comparative images, the high quality of image reconstruction can be directly inferred by simple inspection of Figures 9-11 and Table III. The three processed images are practically undistinguishable from the original, with Pyramid Coding requiring lower bit-rate.

Finally, Figure 12 is "ZELDA" coded with Blockwise Multiresolution scheme in 1.25 b/pel for SNR beyond 38dB threshold.

6) CONCLUSION

We tried to propose Pyramid DPCM as an alternative to conventional DPCM in high sampling rate coding environments. It is more complex, but slower and, since interpolation provides better performance over extrapolation, compression efficiency is improved. One point that must be strongly enphasized is that the results here achieved are too far from optimum. They are relevant when compared with standard DPCM under same conditions, situation into which Pyramid scheme reveals to be an atractive alternative combining Sub-Band, Pyramid and DPCM coding for achieving a superior performance.

The interpolator here used is a discrete extension of the Cubic Convolution Kernel [6] [7] due to its extreme simplicity (even considering its coefficients), adequated polyphase structure [8] and performance. Its impulse response is [7]:

$$h(0)=1 \quad h(\pm1)=9/16 \quad h(\pm2)=0 \quad h(\pm3)=-1/16 \qquad \text{(E8)}$$

$$h_0(n) = \delta(n) \qquad \text{(E9a)}$$

$$h_1(0) = h_1(3) = -1/16\ ;\ h_1(1) = h_1(2) = 9/16 \quad \text{(E9b)}$$

where h(n) is in a non-causal linear-phase presentation and the polyphase filters are in causal format. The filters depicted in Figure 6 are exactly $h_1(n)$ and the equation for interpolation are

$$x_{k-1}(2n+1) = (x_k(n)+x_k(n+1))\frac{9}{16} - (x_k(n-1)+x_k(n+2))\frac{1}{16} + L_{k-1}(n) \qquad \text{(E10)}$$

for k = 1 and 2, n = 0,1...,$N/2^k$ (over one line)

For comparisons, we have done some tests and simulations over 4 test images : "Zelda", "Kitchen" , "Beach Scene" and "Room Scene". In a first step we evaluated the error entropies. Let p(k) be the probability of X=k ($k \in K$); the zero-th order entropies, here considered, are given by :

$$H\ [X] = -\sum_{i \in K} p(i) \log_2 p(i) \qquad \text{(E11)}$$

$$H0 = H\ [x(n)] \qquad \text{(E12a)}$$

$$H1 = H\ [x(n)-x(n-1)] \qquad \text{(E12b)}$$

$$H4 = H\ [x(n)-x(n-4)] \qquad \text{(E12c)}$$

$$H_{L1} = H\ [L_1(n)] \qquad \text{(E12d)}$$

$$H_{L0} = H\ [L_0(n)] \qquad \text{(E12e)}$$

Now, in order to compare the full-image and L_2's DPCM coding for past-sample prediction, we must compare H1 and H4. Furthermore, the entropy gain (G) must also take into account H_{L0} and H_{L1}. In table I, the results of tests over those four images are presented leading to a mean gain around 0.2 b/pel. If prefiltering is allowed, as in TABLE II, the mean gain rises to 0.5 b/pel. This prefiltering will improve interpolation, eliminating aliasing, but it will slightly corrupt the samples in L_2. Repeating the process for planar prediction with $\mathbf{A}$=[1 -0.7 0.7] we have found the mean gain as 0.4 b/pel.

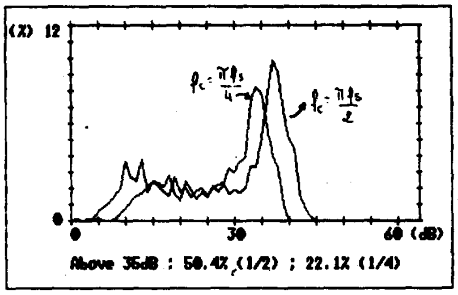
$$G = H1 - \frac{H4+H_{L1}+2H_{L0}}{4} \qquad \text{(E13)}$$

TABLE I Entropies (b/pel) — TABLE II

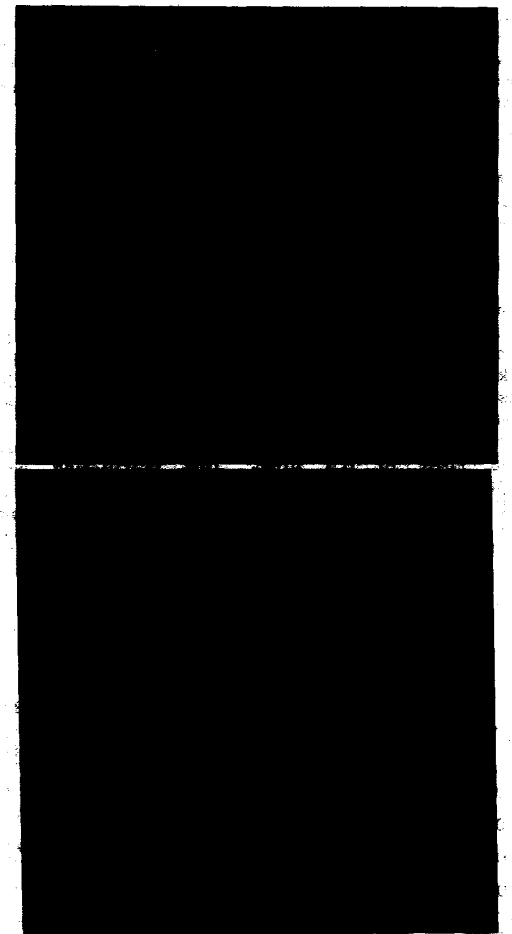
IMAGE	H0	H1	H4	H_{L0}	H_{L1}	G	H4	H_{L0}	H_{L1}	G
BEACH	7.33	4.61	5.97	3.49	4.79	0.18	5.84	2.91	4.45	0.40
ROOM	6.72	3.84	5.05	2.93	3.99	0.12	4.94	2.28	2.22	0.73
ZELDA	6.97	3.40	4.93	2.16	2.81	0.38	4.86	1.50	2.42	0.67
KITCH	6.86	3.56	4.90	2.59	3.56	0.15	4.78	1.98	3.19	0.42

TABLE III Bit-rate(b/pel);SNR(dB)

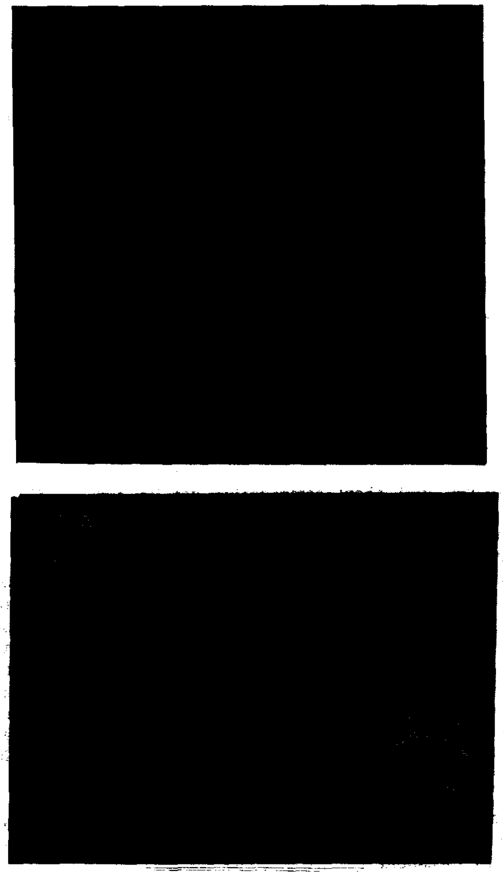
IMAGE	UR PYR 1	BL PYR 2	BR DPCM
ZELDA	1.45/42	2.00/44	2.32/45
KITCH	1.77/43	2.26/44	2.44/45
GIRL	2.45/34	2.80/35	3.05/39



b c
—x—x—x—x—x– field n+1
.....................field n
a
—x—x—o—x—x– field n+1
.....................field n
o – sample to be predicted



Figures 7, 8, 9 and 10



Figures 11 and 12

7) REFERENCES

[1] N.S.Jayant,P.Noll, *Digital Coding of Waveforms,* Englewood Cliffs, NJ, Prentice-Hall, 1984.

[2] P.J.Burt,E.H.Adelson, *The Laplacian Pyramid as Compact Image Code,* IEEE Trans Commun., COM-31, pp 532-540, April 1983.

[3] S.L.Tanimoto,*Image Transmission with Gross Information First,* Comp.Graph and Image Proc. 9, pp 72-76, 1979.

[4] L.Wang,M.Goldberg,*Reduced-difference Pyramid,* Optical Engineering, Vol 28, #7, July 1989.

[5] M.Bellanger, *Adaptive Digital Filters and Signal Analysis,* Marcel Dekker Inc., NY,1987.

[6] R.G.Keys, *Cubic Convolution Interpolation For Digital Image Processing,* IEEE Trans on ASSP, ASSP-29,pp 746-749, June 1983.

[7] R.L.Queiroz,J.B.Yabu-uti, *Técnicas de Projeto de Filtros FIR para Dizimação e Interpolação de Sinais Discretos,* RT-178, Contrato 208/89, UNICAMP/TELEBRÁS, Julho 1989.

[8] R.E.Crochiere,L.R.Rabiner, *Multirate Digital Signal Processing.* Englewood Cliffs, NJ, Prentice-Hall, 1983

[9] M.Hadhoud, D.Thomas, *Two-Dimensional Adaptive LMS Algorithm,* IEEE Trans. on Circuits and Systems, CAS-35, May 1988.