NOISY PCA THEORY AND APPLICATION IN FILTER BANK CODEC DESIGN

Konstantinos I. Diamantaras *Michael G. Strintzis*

Dept. Electrical & Computer Engineering
Aristotle University of Thessaloniki
GR 540 06, Thessaloniki, GREECE

ABSTRACT

Noisy Principal Component Analysis (NPCA) was introduced recently as an extension of PCA in the assumption that the linear features are unreliable. The level of noise in the representation variables is found to have effects in the rank of the optimal solution resembling the water-filling analogy in information theory. The NPCA problem needs to be coupled with certain constraints so that it permits a finite solution. We present the solution of the NPCA problem under different constraints which can be useful in applications involving bandwidth limitations. One of these applications is the design of optimal subband coders incorporating quantization noise. In addition to the NPCA-optimality another advantage of the new design approach is that it works entirely in the time domain and thus the costly and difficult transformations to and from the Z-domain can be avoided.

1. NOISY PRINCIPAL COMPONENT ANALYSIS (NPCA)

Noisy Principal Component Analysis (NPCA) was recently introduced in order to deal with the problem of the optimal linear vector coding and decoding in the presence of noise in the representation variables [3, 7, 1, 4]. Noisy PCA does not require any prior assumptions on the distribution of the signal and can be applied to non-Gaussian signals as well as Gaussian ones. The noise included in the representation or *code vector* $\mathbf{y}$, can be a model of the quantization error in coding applications, of the transmission error in communication applications, of random neural activity in neural network models, etc. A generic diagram describing the NPCA problem is shown in Fig. 1.

Like the standard Principal Component Analysis (PCA) method the dimension m of the *code vector* $\mathbf{y}$, is assumed less than the dimension n, of the input vector $\mathbf{x}$. Unlike PCA however, we assume now that the code vector contains noise uncorrelated with the input, so $E\mathbf{x}\mathbf{e}^T = 0$. Additionally we assume that the input $\mathbf{x}$, is either a stationary signal or a random vector and that the *observation vector* $\hat{\mathbf{x}}$ is degraded by noise ν which is uncorrelated with both $\mathbf{x}$ and $\mathbf{e}$. Analytically, the coding and decoding equations are

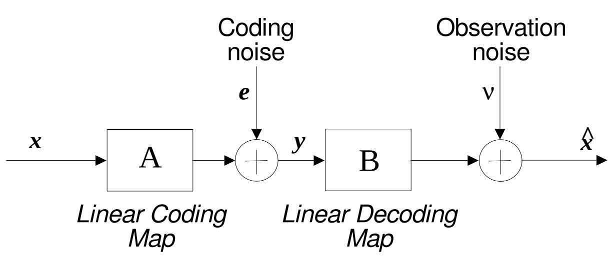


Figure 1. Codec using Principal Component Analysis with noisy representation.

$$\mathbf{y} = \underline{\mathbf{W}}\mathbf{x} + \mathbf{e} \quad (1)$$

$$\hat{\mathbf{x}} = \overline{\mathbf{W}}\mathbf{y} + \nu \quad (2)$$

where $\underline{\mathbf{W}}$, $\overline{\mathbf{W}}$, are the linear coding and decoding operators represented by two matrices.

The usual mean-squared error is minimized for NPCA as with standard PCA:

$$J = E\|\hat{\mathbf{x}} - \mathbf{x}\|^2 \quad (3)$$

It is easy to see that the observation noise ν is inconsequential to the minimization of (3) due to its lack of correlation with either $\mathbf{x}$ or $\mathbf{e}$. Indeed, expanding (3) we obtain

$$\begin{aligned} J &= E\|\mathbf{x} - \overline{\mathbf{W}}\underline{\mathbf{W}}\mathbf{x} - \overline{\mathbf{W}}\mathbf{e}\|^2 \\ &= \operatorname{tr}\{\overline{\mathbf{W}}\underline{\mathbf{W}}\mathbf{R}_x\underline{\mathbf{W}}^T\overline{\mathbf{W}}^T\} - 2\operatorname{tr}\{\overline{\mathbf{W}}\underline{\mathbf{W}}\mathbf{R}_x\} \\ &\quad + \operatorname{tr}\{\overline{\mathbf{W}}\mathbf{R}_e\overline{\mathbf{W}}^T\} + \operatorname{tr}\mathbf{R}_\nu + \operatorname{tr}\mathbf{R}_x \end{aligned} \quad (4)$$

where $\mathbf{R}_x$, $\mathbf{R}_e$, and $\mathbf{R}_\nu$ are the correlation matrices of the signal, the coding noise and the observation noise, respectively. The observation noise enters the equation only through the constant term $\mathbf{R}_\nu$ which can be dropped without affecting the optimal solution.

The coding noise $\mathbf{e}$, on the other hand, can not be ignored and it does affect the optimal solution. Clearly, if the noise power is zero the NPCA problem degenerates into the standard PCA problem. If however, the noise is non-zero then NPCA has many interesting properties. Firstly, there is no finite solution unless we impose some constraint on the matrices $\overline{\mathbf{W}}$ and $\underline{\mathbf{W}}$. Different constraints have been studied in the literature but still many questions are open. Secondly, the problem is closely related with the so-called water-filling analogy appearing in information theory [5]. Thirdly, the rank of the optimal solution under most constraints reduces as the noise power increases.

Next we shall describe the optimal NPCA solutions for various important constraints, although the list is not exhaustive. Some more constraints are discussed for example, in [4].

2. PROBLEM CONSTRAINTS

In the absence of constraints the NPCA problem is unbounded. The infimum is achieved asymptotically for $\underline{\mathbf{W}} \to \infty$ and $\overline{\mathbf{W}} = \underline{\mathbf{W}}^+ \to 0$. With this solution the error term

$\overline{\mathbf{W}}\mathbf{e}$ is diminished relative to the signal term $\overline{\mathbf{W}}\underline{\mathbf{W}}\mathbf{x}$ in the cost function J. In order to obtain a meaningful solution we need to impose an upper limit on the size of the coding matrix $\underline{\mathbf{W}}$ or a lower limit on the size of the decoding matrix $\overline{\mathbf{W}}$.

2.1. Direct coding and decoding matrix constraints

Two of the earlier studied constraints are the straightforward bounds on the Frobenius norms of the coding and decoding matrices [3]:

$$\|\underline{\mathbf{W}}\|_F^2 \leq s^2 \tag{5}$$

$$\|\overline{\mathbf{W}}\|_F^2 \geq s^2 \tag{6}$$

For the NPCA constraints shown above the problem has been solved. Before proceeding to the solution let us introduce some notation: let $(\lambda_i^x, \mathbf{u}_i^x)$, and $(\lambda_i^e, \mathbf{u}_i^e)$ be the eigenvector/eigenvalue pairs of the matrices $\mathbf{R}_x$ and $\mathbf{R}_e$ respectively, such that the eigenvalues are arranged in decreasing order. The following are the basic results [4]:

(a) For the constraint (5) on the coding matrix the optimal solution is

$$\underline{\mathbf{W}} = \sum_{i=1}^{r} \gamma_i \mathbf{u}_{m-i+1}^e \mathbf{u}_i^{x\,T} \tag{7}$$

where the γ_i's are scalar values dependent upon the signal and noise eigenvalues. The optimal rank r is less or equal to m, and it diminishes as the noise power increases. In the limit $\lambda_m^e \to \infty$ we have $r = 1$.

(b) For the constraint (6) on the decoding matrix the optimal solution is not attained for any finite matrix $\overline{\mathbf{W}}$ but we can come arbitrarily close to the optimum for an almost rank-1 matrix

$$\overline{\mathbf{W}} = s\mathbf{u}_1^x \mathbf{u}_m^{e\,T} + \sum_{i=2}^{m} \epsilon \mathbf{u}_i^x \mathbf{u}_{m-i+1}^{e\,T} \tag{8}$$

where *epsilon* is an arbitrarily small positive constant.

What is common in both cases is the low-rank property of the optimal solution. Especially in the first case the optimal rank reduces as the noise power increases. This has been likened to the water-filling analogy found in the information theoretical context when a signal has to pass through a number of parallel Gaussian channels with different noise levels. See [5] for more details. We must note that this pattern of rank-reduction in relation with the increase of the noise power appears in the solution of various other NPCA constraints as well.

2.2. Representation variance constraint

In most applications channel bandwidth is at a premium. In these cases it is reasonable to put an upper bound on the NPCA code vector variance thus indirectly limiting the size of coding matrix $\underline{\mathbf{W}}$ as well. Consider the constraint

$$E\|\mathbf{y}\|^2 = s^2 \tag{9}$$

for some fixed value s. Using the method of Lagrange multipliers for optimizing (3) under the constraint (9) we define the auxiliary cost $J' = J + \mu(E\|\mathbf{y}\|^2 - s^2)$. According to Theorem 5.5 in [4], for fixed μ the optimal solution is

$$\underline{\mathbf{W}} = \sum_{i=1}^{m} \underline{\sigma}_i \mathbf{u}_{m-i+1}^e \mathbf{u}_i^{x\,T} \qquad \overline{\mathbf{W}} = \sum_{i=1}^{m} \overline{\sigma}_i \mathbf{u}_i^x \mathbf{u}_{m-i+1}^{e\,T} \tag{10}$$

$$\underline{\sigma}_i^2 = \begin{cases} 1/\sqrt{\gamma_i \mu} - 1/\gamma_i & \text{if } \mu < \gamma_i, \\ 0 & \text{otherwise} \end{cases}$$

$$\overline{\sigma}_i^2 = \begin{cases} \sqrt{\gamma_i \mu} - \mu & \text{if } \mu < \gamma_i, \\ 0 & \text{otherwise} \end{cases}$$

where $\gamma_i = \lambda_i^x / \lambda_{m-i+1}^e$ are signal to noise ratios with the signal eigenvalues paired in reverse order with the noise eigenvalues (strongest signal component with weekest noise component).

For the special case where the noise components are uncorrelated and have the same variance λ^e, the eigenvectors $\mathbf{u}_i^e$ can be any set of orthonormal vectors. For this case the variance of the noisy representation is

$$E\|\mathbf{y}\|^2 = \sqrt{\lambda^e}/\sqrt{\mu} \sum_{i=1}^{r} \sqrt{\lambda_i^x} + (m-r)\lambda^e \tag{11}$$

where $r \leq m$ is the rank of the optimal solution. The variance constraint (9) in this case gives us the optimal value of μ

$$\mu = \frac{\lambda^e [\sum_{i=1}^r \sqrt{\lambda_i^x}]^2}{[s^2 - (m-r)\lambda^e]^2} \tag{12}$$

After some mathematical manipulations we find that the rank r is a function of the representation variance s and is given by the relations

$$\begin{array}{lcl} s^2 = l(1) = m\lambda^e & \Rightarrow & r = 0 \\ l(1) < s^2 \leq l(2) & \Rightarrow & r = 1 \\ l(2) < s^2 \leq l(3) & \Rightarrow & r = 2 \\ & \vdots & \\ l(m) < s^2 & \Rightarrow & r = m \end{array} \tag{13}$$

where $l(r) = \lambda^e \sum_{i=1}^r \sqrt{\lambda_i^x / \lambda_r^x} + (m-r)\lambda^e$.

3. FILTER BANK DESIGN

Noisy PCA can be applied in designing optimal subband coding systems incorporating quantizers. These coders comprise analysis / synthesis filters which split the signal into frequency bands and code each band separately using a quantizer optimized for the band's statistics. Subband coding techniques have become popular recently due to their superior performance in many applications such as image coding. The design of subband coding systems is typically done in the frequency domain requiring the computation of the input signal power spectrum $\Phi_x(z)$. Until recently most subband coders were designed ignoring the error introduced by the quantizers used in each subband. A total system design including the quantization error was studied in [11, 9, 6]. Next, we shall show how to formulate the design problem of an optimal subband coder as a Noisy PCA problem. The computational advantages of this approach stem from the fact the Noisy PCA is a pure time-domain technique and the costly transformations to and from the frequency domain can be avoided. Furthermore, there is an adaptive approach for estimating the Noisy PCA solution in environments where the exact signal statistics are unknown.

Consider the generic M-bank filter shown in Fig. 2. The analysis stage (Fig. 2a) has a typical filtering-followed-by-subsampling structure. Here we assume that each filter H_i is FIR with order $L = M$. Let the input samples $x(kM)$,

$x(kM-1)$, ..., $x(kM-M+1)$, form a vector $\mathbf{x}(k)$. Then the variables v_i result from a linear operation

$$\mathbf{v}(k) = [v_0(k)\, v_1(k)\, \cdots\, v_{M-1}(k)]^T = \mathbf{H}\mathbf{x}(k) \tag{14}$$

where the elements of matrix $\mathbf{H}$ are the taps $h_{i,n}$, $n = 0, 1, \cdots, n$, of the filters H_i: $\mathbf{H}_{ij} = H_{i-1,j-1}$.

For the quantization of the subbands v_0, ..., v_{M-1}, we assume the use of optimal Lloyd-Max quantizers designed according to the statistics of each subband. The Lloyd-Max quantizers are modeled using a gain parameter α, and an additive noise term e, as follows [8]

$$y = \alpha v + e \tag{15}$$

The gain parameter depends on the input signal and its value is determined by the formula

$$\alpha = 1 - \sigma_q^2/\sigma_v^2 \tag{16}$$

where σ_q^2 is the quantization noise variance. Using the value (16) for α the additive noise component e is uncorrelated with the quantizer input v. It can be shown that the variances of the input, of the output, and of the quantization error are related by the formula $\sigma_y^2 = \sigma_v^2 - \sigma_q^2$, so from (15) and (16) we have $\sigma_e^2 = \alpha(1-\alpha)\sigma_v^2$. If we define $\tilde{v} = \alpha v$, then $\sigma_{\tilde{v}}^2 = \alpha^2\sigma_v^2$, so

$$\sigma_e^2 = (\frac{1}{\alpha} - 1)\sigma_{\tilde{v}}^2 \tag{17}$$

From rate distortion theory [2] it can be shown that the parameter α is related with the number of bits R used for coding v by the formula $\alpha = 1 - \beta(R)2^{-2R}$ where β is an increasing function of R with minimum value $\beta(0) = 1$ and a maximum asymptotic value $\beta_\infty > 1$.

Using (15) the analysis plus quantization stage can be represented by the equations

$$\tilde{\mathbf{v}}(k) = [\tilde{v}_0(k)\, \tilde{v}_1(k)\, \cdots\, \tilde{v}_{M-1}(k)]^T = \mathbf{A}\mathbf{H}\mathbf{x}(k) \tag{18}$$

where $\mathbf{A} = \mathrm{diag}[\alpha_0\, \alpha_1\, \cdots\, \alpha_{M-1}]$.

In the synthesis stage (Fig. 2b) the upsampling is followed by the synthesis filter bank. Assuming that the filters G_i are all FIR of order $L = M$, it is easy to verify that the following equation holds

$$\hat{x}(kM+n) = \sum_{i=0}^{M-1} g_{i,n} y_i(k), \qquad n = 0, 1, \cdots, M-1 \tag{19}$$

or

$$\hat{\mathbf{x}}(k) = [\hat{x}(kM+M-1)\, \cdots\, \hat{x}(kM+1)\, \hat{x}(kM)]^T = \mathbf{G}\mathbf{y}(k) \tag{20}$$

where $g_{i,n}$, $n = 0, 1, \cdots, M-1$, are the taps of filter G_i, $\mathbf{y}(k) = [y_0(k)\, y_1(k)\, \cdots\, y_{M-1}(k)]^T$, and $[\mathbf{G}]_{ij} = g_{M-i,j-1}$.

Using (18) and (20) and introducing the noise vector $\mathbf{e} = [e_0(k)\, e_1(k)\, \cdots e_{M-1}(k)]^T$ we obtain the equations describing the overall subband coder system

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{A}\mathbf{H}\mathbf{x}(k) + \mathbf{e}(k) & (21)\\ \hat{\mathbf{x}}(k) &= \mathbf{G}\mathbf{y}(k) & (22) \end{aligned}$$

Assuming that $\hat{x}(k)$ be a (delayed) estimate of $x(k-M+1)$ then we want to minimize $J = E\|\mathbf{x} - \hat{\mathbf{x}}\|^2$.

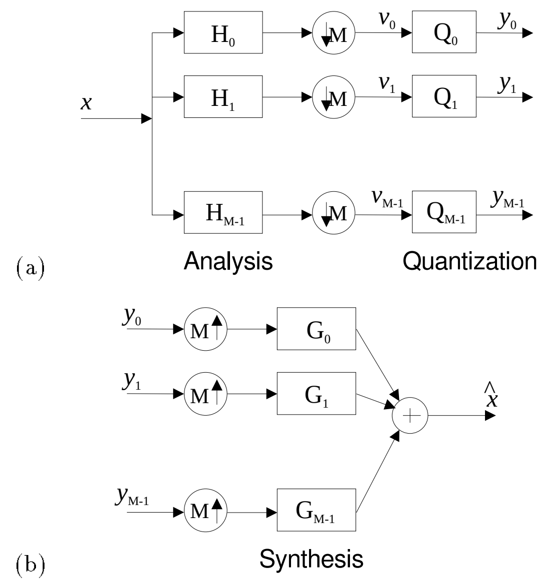


Figure 2. Typical Filter Bank (a) Analysis stage, and (b) Synthesis stage.

Lloyd-Max quantizers are optimized for the statistics of each variable v_i. Consequently, we assume that the noises e_i are uncorrelated with the variables v_i, with the signal x, and also with each other. Thus the noise covariance matrix is diagonal $\mathbf{R}_e = \mathrm{diag}[\sigma_{e_1}^2\, \sigma_{e_2}^2\, \cdots\, \sigma_{e_{M-1}}^2]$ and the noise eigenvalues are equal to the individual noise variances.

The bandwidth of the channel through which we transmit the quantized values y_i in addition to the compression and coding methods used for the transmission set an upper limit s^2 on the variance of the transmitted signal. Using a maximum variance constraint such as $E\|\mathbf{y}\|^2 = s^2$, we derive the optimal analysis and synthesis filters using the Noisy PCA approach as follows:

1. Select $\sigma_{q_0}^2 = \cdots = \sigma_{q_{M-1}}^2 = \sigma_q^2$, and select μ.
2. Compute
$$\lambda_{m-i+1}^e = \frac{1}{4}\left[\sqrt{\frac{\mu}{\lambda_i^x}} + \sqrt{\frac{\mu}{\lambda_i^x}\sigma_q^4 + 4\sigma_q^2}\right]^2$$
3. Determine optimal rank from the conditions $\mu <=> \lambda_i^x/\lambda_{m-i+1}^e$
4. Compute
$$s^2 = \frac{1}{\sqrt{\mu}}\sum_{i=1}^{r}\sqrt{\lambda_{m-i+1}^e\lambda_i^x} + \sum_{i=r+1}^{m}\lambda_{m-i+1}^e$$
If s^2 is satisfactory, goto 4. If s^2 is too large increase μ goto 2; if s^2 is too small decrease μ goto 2.
5. Compute NPCA solution $\underline{\mathbf{W}}$, $\overline{\mathbf{W}}$, and let
$$\alpha_i = 1 - \sqrt{\mu\lambda_{m-i+1}^e/\lambda_i^x}$$
6. Obtain optimal filters $\mathbf{H} = \mathbf{A}^{-1}\underline{\mathbf{W}}$, $\mathbf{G} = \overline{\mathbf{W}}$.

The proposed method is detailed in the following example where we use Noisy PCA to design an optimal subband coder for an image.

4. APPLICATION EXAMPLE: IMAGE CODING

We use a large number of images in order to statistically estimate the image covariance function. Under the usual Markovian assumption for the image probability density, the cross-covariance for two pixels with displacement Δx, Δy, is modeled as [10]

$$r(\Delta x, \Delta y) = \sigma \rho_x^{|\Delta x|} \rho_y^{|\Delta y|}$$

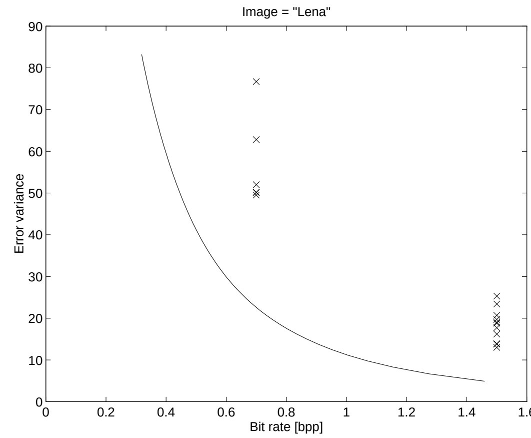
The parameters were estimated as $\sigma = 2693$, $\rho_x = 0.923$, and $\rho_y = 0.977$.

Next we form the signal correlation matrix $\mathbf{R}_x$ of the input vectors which are 8×8 image blocks. The eigenvalue decomposition of $\mathbf{R}_x$ yields the eigenvalues λ_i^x and the eigenvectors $\mathbf{u}_i^x$. The noise covariance matrix is assumed diagonal. We follow the procedure described in the previous Section in order to design the NPCA-optimal 64-bank filter. The rate-distortion plots for the image "Lena" are shown in Figure 3a. μ takes a range of values from 10 to 500 while σ_q^2 is fixed at 10^{-4}. The results are compared with the ones obtained in [9] for different filters and for the bit rates 0.7bpp and 1.5bpp. Figure 3b shows the optimal solution rank. Observe that even for 1.5bpp the number of components used are 45 out 64. For rate 0.5bpp approximately 15 components are used out of 64.

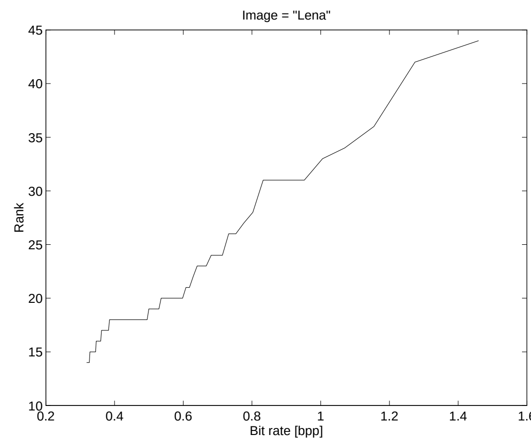
5. CONCLUSIONS

The novel Noisy PCA method deals with unreliability in the linear feature extraction process. It shares some common properties with standard PCA, such as the involvement of the eigenvalue decomposition of the signal correlation matrix. It also has some distinct differences from PCA notably, the low-rankness of the optimal solution and the absence of a finite solution without imposing any constraints. Preliminary theoretical results have been derived for different constraints, some of which can have application in signal/image processing problems.

The design of filter bank coding systems incorporating the noise from the quantization operations is one of the application areas where a specific NPCA constraint can be used. The results on standard images show promise for a more general application of NPCA concepts in image coding.



(a)



(b)

Figure 3. (a) Rate-distortion plot for the NPCA-optimal 64-bank filter using the 256×256 image "Lena". As a comparison we show the corresponding R-D points obtained in [9] for different filters (points marked by 'x'). (b) The optimal rank as a function of the bit-rate.

REFERENCES

[1] P. Baldi and K. Hornik. "Learning in Linear Neural Networks: a Survey". *IEEE Trans. Neural Networks*, vol. 6, no. 4, pp. 837–858, July 1995.

[2] T. Berger, *Rate Distortion Theory.* Prentice Hall, NJ, 1971.

[3] K. I. Diamantaras and K. Hornik. "Noisy Principal Component Analysis". In J. Volaufova and V. Witkowsky, editors, *Measurement '93*, pages 25–33, Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia, May 1993.

[4] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications.* John Wiley, New York, 1996.

[5] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, NY, 1968.

[6] R. A. Haddad and K. Park, "Modeling, Analysis, and Optimum Design of Quantized M-Band Filter Banks". IEEE Trans. Signal Processing, pp. 2540–2549, vol. 43, no. 11, Nov. 1995.

[7] K. Hornik, "Noisy Linear Networks". In R. Mammone, editor, in *Artificial Neural Networks with Applications in Speech and Vision*, Chapman & Hall, London, 1994.

[8] N. S. Jayant and P. Noll, *Digital Coding of Waveforms.* Prentice-Hall, NJ, 1984.

[9] J. Kovačević, "Subband Coding Systems Incorporating Quantizer Models". IEEE Trans. Image Processing, pp. 543–553, vol. 4, no. 5, May 1995.

[10] A. N. Netravali and B. G. Haskell, *Digital Pictures, Representation and Compression*, Plenum Press, 1988.

[11] P. Westerink, J. Biemond, and D. Boekee, "Scalar Quantization Error Analysis for Image Subband Coding Using QMF's". Signal Processing, pp. 421–428, vol. 40, Feb. 1992.