MDC image coding using Cascaded Correlating Transforms

Ruediger Knoerig and Thomas Sikora

University of Technology Berlin, Department for telecommunication systems and signal processing, Einsteinufer 17, D-10587 Berlin, Germany

ABSTRACT

This paper describes a joint source-channel coding framework combining cascaded correlating transforms as proposed by Goyal in[1] with an optimal estimation algorithm in the MSE sense. The cascaded correlating transform, an extension of the well-known (see [2], [3], [4] or [5] for example) pairwise correlating transform to transforms of higher order, can be seen as a detunable decorrelating transform. By reducing the transforms ability to decorrelate, a higher amount of source correlation "survives" in the signal. This increased redundancy will be used for concealing channel errors.

Since the detuning can be performed stepless an arbitrary amount of redundancy can be selected, allowing fine-tuned trade-offs between coding efficiency and robustness to channel errors. This is an advantage over the classic approach by combining source- and channel coders since even shortened convolution coders offer only a discrete and therefore not stepless set of coding rates. Moreover, our approach affects only the transform and the inverse transform stages and will be transparent to other stages of the coding system (e.g. quantization or entropy coding).

1. INTRODUCTION

The basic concept of our proposed framework is depicted in figure 1. It shows a classic multiple description transform coding framework. On the left we start by sequencing the input data into blocks (not shown) and transform these blocks $\mathbf{x}[n] = (x_0[n], x_1[n], \ldots, x_{M-1}[n])$ into the transform domain representation $\mathbf{y}[n] = (y_0, y_1, \ldots, y_{M-1})$. Then each transform coefficient $y_k[n]$ will be coded and transmitted independently, following the multiple description coding approach. On the receiver side we have to deal with the situation that there may be only a subset of the transmissions available since some of the transmissions may be corrupted by the faulty channel or will arrive too late for being useful in streaming applications. Therefore we should have an inverse transform which makes the best out of even a subset of transform coefficients.

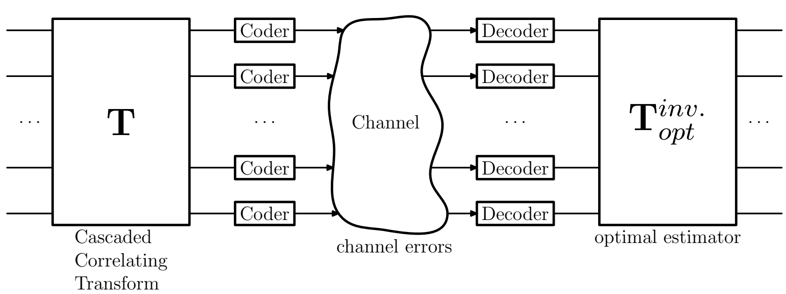

Figure 1. The building blocks of our framework.

Next we will describe the most important building blocks of our framework, the cascaded correlating transform stage and the optimal estimator used for the inverse transformation. Further we will present the results of our image coding experiments.

[knoerig | sikora]@nue.tu-berlin.de

2. THE CASCADED CORRELATION TRANSFORM.

In[1] Goyal proposed the cascading of the 2x2 pairwise correlating transforms

$$\mathbf{T}_\alpha = \begin{pmatrix} \alpha & 1/2\alpha \\ -\alpha & 1/2\alpha \end{pmatrix} \text{ and } \mathbf{T}_\alpha^{-1} = \begin{pmatrix} \frac{1}{2\alpha} & -\frac{1}{2\alpha} \\ \alpha & \alpha \end{pmatrix} \tag{1}$$

to yield correlating transforms of higher order. Notice that the determinant of the transform matrix always equals one. Therefore the inverse transform exists for all values of α. The principle of the correlating transform

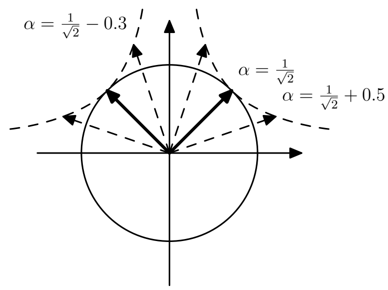

Figure 2. Base functions of the pairwise transform with respect to α

reveals itself if we have a look at its base vectors with respect to the parameter α, as depicted in figure 2. As one can see, we get an orthonormal, Haar-like base for $\alpha = \frac{1}{\sqrt{2}}$. Other values $\alpha = \frac{1}{\sqrt{2}} + \varepsilon$ results in more or less nonorthogonal bases. Therefore an $\varepsilon \neq 0$ "detunes" the basically decorrelating transform.

This becomes more obvious if we have a look at the exact value of the normalized correlation between the transform coefficients, which can be calculated from the transform coefficients easily. The transform equation

$$\vec{y} = \mathbf{T}_\alpha \cdot \vec{x} \tag{2}$$

provides the transform coefficients as

$$y_0 = \alpha \cdot x_0 + \frac{x_1}{2\alpha} \tag{3}$$

$$y_1 = -\alpha \cdot x_0 + \frac{x_1}{2\alpha}\,. \tag{4}$$

Under the assumption* $\sigma_{x_1}^2 = \sigma_{x_2}^2$ the normalized correlation between the transform coefficients is simply

$$\rho_{out} = \frac{E\{y_0 \cdot y_1\}}{\sqrt{\sigma_{y_1}^2 \cdot \sigma_{y_2}^2}} \tag{5}$$

$$= \frac{\mathbf{R}_{yy}(1,1)}{\sqrt{\mathbf{R}_{yy}(1,1) \cdot \mathbf{R}_{yy}(2,2)}} \tag{6}$$

with

$$\mathbf{R}_{yy} = E\{y \cdot y^H\} \tag{7}$$

$$= \sigma_x^2 \cdot \begin{pmatrix} \frac{1}{4\alpha^2} + \alpha^2 + \rho_{01}^{(x)} & \frac{1}{4\alpha^2} - \alpha^2 \\ \frac{1}{4\alpha^2} - \alpha^2 & \frac{1}{4\alpha^2} + \alpha^2 - \rho_{01}^{(x)} \end{pmatrix} \tag{8}$$

$$\Rightarrow \rho_{out} = \frac{\frac{1}{4\alpha^2} - \alpha^2}{\sqrt{\left(\frac{1}{4\alpha^2} + \alpha^2\right)^2 - \rho_{in}^2}} \tag{9}$$

*which holds for the transform coding purpose since there should no variance differences between even and odd source samples

Equation 9 shows that the correlation will be zero for the orthogonal case $\varepsilon = 0$ and will raise if α strays further away from $1/\sqrt{2}$. The disadvantage of the gain in correlation is the degradation of the transform coding gain G_{TC} (see[6])

$$G_{TC} = \frac{\frac{1}{N}\sum_{k=0}^{N-1}\sigma_{y_k}^2}{\left(\prod_{k=0}^{N-1}\sigma_{y_k}^2\right)^{\frac{1}{N}}} \tag{10}$$

$$= \frac{\frac{1}{2}\cdot\left(\sigma_{y_0}^2+\sigma_{y_1}^2\right)}{\sqrt{\sigma_{y_0}^2\cdot\sigma_{y_1}^2}} \tag{11}$$

$$= \frac{\frac{1}{4\alpha^2}+\alpha^2}{\sqrt{\left(\frac{1}{4\alpha^2}+\alpha^2\right)^2-\rho_{in}^2}} \tag{12}$$

Thus we're able to seek a fine-tuned trade-off between coding efficiency and additional redundancy for the purpose of error concealment by varying the parameter ε.

2.1. Cascaded transforms

The cascading structure proposed by Goyal is depicted in figure 3. Its easy to show that the cascading structure in figure 3 for correlating transforms as suggested in[1] equals a subband coder as depicted in figure 4. The first stage of the subband coder transforms two adjacent input samples $x[n]$ and $x[n-1]$ into the transform coefficients $t_0[n]$ and $t_1[n]$. The two analysis filter banks in the second stage transform two adjacent output samples of the first stage, $t_{0/1}[n]$ and $t_{0/1}[n-1]$, to two transform coefficients $y_{0/2}[n]$ and $y_{1/3}[n]$.

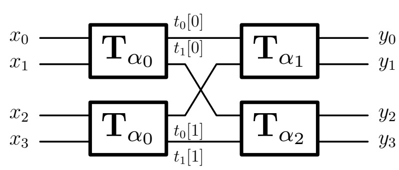

Figure 3. Cascaded correlating transforms as suggested by.[1]

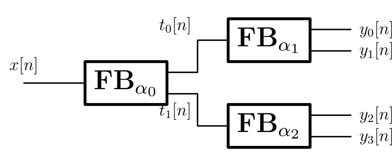

Figure 4. Cascaded filterbanks for subband coding with $M = 4$.

The only difference to the cascaded correlating transforms in figure 3 is that the processing of two times two adjacent input samples $((x[n-3]\ x[n-2])$ and $(x[n-1]\ x[n]))$ is not done sequentially but in parallel by two instead of one transformers. Thus both viewpoints describe the same transform. The interpretation of the cascade as a subband coder is the key for understanding the multiresolutional way in which the redundancy will be added.

As long as all cascaded elementary transforms uses the same value for α the rule for calculating the corresponding transform matrices is simply

$$\mathbf{T}_{M=4} = \mathbf{T}_{M=2} \otimes \mathbf{T}_{M=2} \tag{13}$$

where $\otimes$ denotes Kronecker multiplication and $\mathbf{T}_{M=2}$ is given by equation 1. Thus for the case $M = 4$ we get

$$= \begin{pmatrix} \alpha^2 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4\alpha^2} \\ -\alpha^2 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{4\alpha^2} \\ -\alpha^2 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4\alpha^2} \\ \alpha^2 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{4\alpha^2} \end{pmatrix} \tag{14}$$

This holds also for the inverse transform:

$$\mathbf{T}^{-1}_{M=4} = \mathbf{T}^{-1}_{M=2} \otimes \mathbf{T}^{-1}_{M=2} \tag{15}$$

$$= \begin{pmatrix} \frac{1}{4\alpha^2} & -\frac{1}{4\alpha^2} & -\frac{1}{4\alpha^2} & \frac{1}{4\alpha^2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \alpha^2 & \alpha^2 & \alpha^2 & \alpha^2 \end{pmatrix} \tag{16}$$

We can extend this pattern recursively and get

$$\mathbf{T}_{M=2N} = \mathbf{T}_{M=N} \otimes \mathbf{T}_{M=N} \tag{17}$$

$$\mathbf{T}^{-1}_{M=2N} = \mathbf{T}^{-1}_{M=N} \otimes \mathbf{T}^{-1}_{M=N} \tag{18}$$

The resulting transform will be called CCT (**C***ascaded* **C***orrelating* **T***ransform*) from now.

3. DERIVATION OF THE ESTIMATOR

3.1. Transform coding basics & terminology

In transform coding we map an input vector $\mathbf{x}$ onto a transformed vector $\mathbf{y}$ via a linear transformation conducted by multiplication with a matrix $\mathbf{T}$.

$$\mathbf{y} = \mathbf{T} \cdot \mathbf{x} \tag{19}$$

Each component y_l of $\mathbf{y}$ is the inner product of the input vector $\mathbf{x}$ and the corresponding row vector of the transform matrix $\mathbf{\Theta}_l$

$$y_l = \mathbf{\Theta}_l \cdot \mathbf{x} \tag{20}$$

The inverse transform is conducted by the inverse transformation matrix $\mathbf{T}^{-1}$

$$\mathbf{y} = \mathbf{T}^{-1} \cdot \mathbf{x} \tag{21}$$

Equation 21 describes the input vector $\mathbf{x}$ as a weighted sum of the inverse transform matrix's column vectors $\tilde{\mathbf{\Theta}}_l$ where the weights are the transform coefficients

$$\mathbf{x} = \sum_{y_l \in \Omega} \tilde{\mathbf{\Theta}}_l \cdot y_l \tag{22}$$

Now we can state our problem as the quest for the optimal estimation matrix $\tilde{\mathbf{T}}$ which estimates $\mathbf{x}$ from an arbitrary subset $S \subset \Omega$ from the set $\Omega = \{y_0, y_1, \ldots, y_{M-1}\}$ of all transform coefficients y_l

$$\tilde{\mathbf{x}} = \sum_{y_l \in S} \tilde{\mathbf{\Theta}}_l \cdot y_l \tag{23}$$

such that the mean square error of the estimation

$$MSE = E\left\{|\mathbf{x} - \tilde{\mathbf{x}}|^2\right\} \tag{24}$$

becomes minimal. Therefore we have to solve

$$\frac{d}{d\tilde{\mathbf{\Theta}}_l} E\left\{\left(\mathbf{x} - \sum_{k\in S} \tilde{\mathbf{\Theta}}_k \cdot y_k\right)^2\right\} \stackrel{!}{=} 0 \tag{25}$$

which turns to

$$E\left\{\left(\mathbf{x} - \sum_{k\in S} \tilde{\mathbf{\Theta}}_k \cdot y_k\right) \cdot y_l\right\} \stackrel{!}{=} 0 \tag{26}$$

$$E\left\{y_l \cdot \mathbf{x})\right\} = \sum_{k\in S} \tilde{\mathbf{\Theta}}_k \cdot r_{kl}^{(y)} \tag{27}$$

after we transpose bot sides we get

$$E\left\{y_l \cdot \mathbf{x}^H)\right\} = \sum_{k\in S} \tilde{\mathbf{\Theta}}_k^H \cdot r_{kl}^{(y)} \,. \tag{28}$$

If we use the transform equation $y_l = \mathbf{\Theta}_l \cdot \mathbf{x}$ the problem becomes

$$E\left\{\mathbf{\Theta}_l \cdot \mathbf{x} \cdot \mathbf{x}^H)\right\} = \sum_{k\neq k_{missing}} \tilde{\mathbf{\Theta}}_k^H \cdot r_{kl}^{(y)} \tag{29}$$

and finally

$$\mathbf{\Theta}_l \cdot \mathbf{R}_{xx} = \sum_{k\in S} \tilde{\mathbf{\Theta}}_k^H \cdot r_{kl}^{(y)} \,. \tag{30}$$

To simplify the following equations we define two matrices. The first one, $\mathbf{T}_{red.}$, is defined as

$$\{\mathbf{\Theta}_l\}_{l\in S} = \mathbf{T}_{red.} \tag{31}$$

and is simply the transform matrix $\mathbf{T}$ where all rows belonging to missing coefficients have been erased. Now the left side of equation 30 becomes

$$\{\mathbf{\Theta}_l \cdot \mathbf{R}_{xx}\}_{l\in S} = \mathbf{T}_{red.} \cdot \mathbf{R}_{xx} \,. \tag{32}$$

The right side of equation 30 can be expressed in the same way

$$\sum_{k\in S} \tilde{\mathbf{\Theta}}_k^H \cdot r_{kl}^{(y)} = \mathbf{R}_{yy_{red.}} \cdot \tilde{\mathbf{T}}_{inv.}^H \tag{33}$$

where $\mathbf{R}_{yy_{red.}}$ is $\mathbf{R}_{yy}$ where all rows *and* columns belonging to the set of missing coefficients $M = \Omega - S$ have been deleted. Equation 30 therefore becomes

$$\mathbf{T}_{red.} \cdot \mathbf{R}_{xx} = \mathbf{R}_{yy_{red.}} \cdot \tilde{\mathbf{T}}_{inv.}^H \,. \tag{34}$$

Now we can solve for the desired optimal $\tilde{\mathbf{T}}_{inv.}$:

$$\tilde{\mathbf{T}}_{inv.} = \left(\mathbf{R}_{yy_{red.}}^{-1} \cdot \mathbf{T}_{red.} \cdot \mathbf{R}_{xx}\right)^H \,. \tag{35}$$

This solution is identical to the optimal solution given in[7] since

$$\Lambda_X = T_{red.} \cdot \mathbf{R}_{xx} \cdot T_{red.}^H \hat{=} R_{yy_{red.}} \tag{36}$$

and

$$\mathbf{R}_{x,Xcenter} = \mathbf{R}_{xx} \cdot T_{red.}^H \tag{37}$$

if we neglect the quantization effect given as noise matrix $\mathbf{W}_X$ there.

Equation 30 has an unique solution if we assume an optimally decorrelating transform. Then the cross-correlation of the transform coefficients becomes

$$r_{kl}^{(y)} = \delta(k-l) \cdot \sigma_{y_l}^2 \tag{38}$$

and equation 30 therefore

$$\mathbf{\Theta}_l \cdot \mathbf{R}_{xx} = \tilde{\mathbf{\Theta}}_l^H \cdot \sigma_{y_l}^2 \tag{39}$$

$$\Rightarrow \tilde{\mathbf{\Theta}}_l^H = \mathbf{\Theta}_l \cdot \mathbf{R}_{xx} \cdot \frac{1}{\sigma_{y_l}^2} \tag{40}$$

The transform base vectors of the optimally decorrelating transform, the Karhunen-Loeve transform (see[6] for details), are the transposed eigenvectors v_i^H of the correlation matrix of the source

$$\{\mathbf{\Theta}\} = eig\left(\mathbf{R}_{xx}\right)^H \tag{41}$$

and the corresponding eigenvalues are the variances of the transform coefficients

$$\lambda_k = \sigma_{y_k}^2 \ . \tag{42}$$

If we apply the eigenvalue problem

$$\mathbf{A} \cdot \lambda_i = \mathbf{v}_i \cdot \lambda_i \tag{43}$$

to equation 40 we see that with the symmetry property of $\mathbf{R}_{xx}$ in mind

$$\tilde{\mathbf{\Theta}}_l \cdot \sigma_{y_l}^2 = \left(\mathbf{\Theta}_l \cdot \mathbf{R}_{xx}\right)^H = \mathbf{R}_{xx} \cdot \mathbf{\Theta}_l^H \stackrel{!}{=} \left(\lambda_l \cdot \mathbf{v}_l\right)^H \tag{44}$$

$$\Rightarrow \tilde{\mathbf{\Theta}}_l = \sigma_{y_l}^2 \cdot \mathbf{\Theta}_l \cdot \frac{1}{\sigma_{y_l}^2} \tag{45}$$

$$= \mathbf{\Theta}_l^H \tag{46}$$

Thus we get the well-known solution that for the optimal decorrelating transform, the KLT, the best estimation is simply the inverse transform itself.

4. MDC IMAGE CODING

For evaluating the performance and usefulness we wrote a demonstrator/evaluation application which performs a simulated MDC image transmission over faulty Gilbert-Elliot channels. You may verify our results by yourself if you download this application from our webpage at `http://www.nue.tu-berlin.de/wer/knoerig/javaprogs.html#DemonstratorOptEst`.

The demonstrator subdivides the image into 8×8 macroblocks and strings together the row vectors of each macroblock to a vector of dimension 64. Then these vectors will be transformed with the selected transform – currently implemented are the DCT, the KLT, the identity transform and of course the CCT. The entropie of the transform coefficients will be estimated via a histogram analysis. Afterwards the transmission of each transform coefficients over 64 independent channels will be simulated using error patterns generated by individual gilbert-elliot channels. For comparing purposes its possible to reuse the error patterns. On the receiver side you can choose if the normal inverse transform should be used or the derived global optimal estimator solution following equation 30.

Transform	SNR_{dB}	$pSNR_{dB}$	Entropie
DCT	20.16 dB	1.67 dB	5.44 Bit/Pixel
DCT w. estimator	20.27 dB	1.78 dB	5.44 Bit/Pixel
KLT	22.17 dB	3.66 dB	4.75 Bit/Pixel
KLT w. estimator	22.17 dB	3.66 dB	4.75 Bit/Pixel
Identity	20.22 dB	1.73 dB	7.26 Bit/Pixel
Identity w. est.	53.68 dB	35.15 dB	7.26 Bit/Pixel
$CCT_{\varepsilon=0.0}$	20.17 dB	1.68 dB	5.26 Bit/Pixel
$CCT_{\varepsilon=0.0}$ + g.e.	20.33 dB	1.83 dB	5.26 Bit/Pixel
$CCT_{\varepsilon=0.02}$	20.10 dB	1.60 dB	5.35 Bit/Pixel
$CCT_{\varepsilon=0.02}$ + g.e.	38.35 dB	19.81 dB	5.35 Bit/Pixel
$CCT_{\varepsilon=0.08}$	19.04 dB	0.54 dB	5.81 Bit/Pixel
$CCT_{\varepsilon=0.08}$ + g.e.	47.63 dB	29.10 dB	5.81 Bit/Pixel

Table 1. Numeric simulation results (from the picture examples).

4.1. Results

Next we show the results from simulated transmissions of the well-known Lena image over independent Gilbert-Elliot channels, each with a packet error rate of $P_{err} = 10^{-2}$ and an average burst length of $L_{burst} = 5$. We reused the error patterns for each simulation.

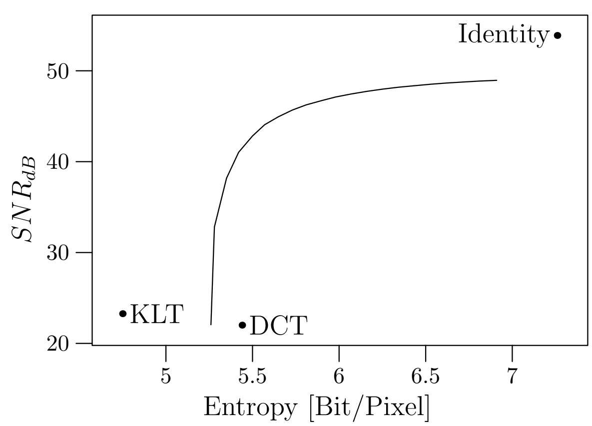

Figure 5. Rate-distortion development for the CCT in combination with the globally optimal estimation.

As one can clearly see from the results in the figures 6-10, the global optimal estimator yields at best for the nonorthogonal transforms. Table 1 shows the dramatic increase of both SNR and $pSNR$ values in case of the nonorthogonal variants of the CCT in combination with the global estimator.

Compared to the orthogonal variant of the CCT, the CCT with $\varepsilon = 0.02$ results in a SNR gain of 18.18 dB for the cost of a very slight increase of the entropy by 0.09 Bit/Pixel. It should be noticed that for this image the CCT with $\varepsilon = 0.02$ has a lower entropie rate than the DCT (5.35 to 5.44 Bit/Pixel).

If we increase ε further to 0.08 we get a SNR gain of 27.46 dB paid with a rate increase of only 0.55 Bit/Pixel. As the pictures 9 and 10 shows these gains are not only of theoretic matter.

More surprising are the small but existent gains for the orthogonal transforms DCT and $CCT_{\varepsilon=0.0}$. The gains are small (0.11 dB resp. 0.16 dB) but this clearly shows that these transforms don't decorrelate completely. In accordance to equation 46 the globally optimal estimator doesn't provide any gain for the optimal decorrelating transform, the KLT.

The result for the identity transform in the figure 7 shows that our estimation algorithm may be useful for replacing pixels in images known as wrong.

Figure 6. Result for the DCT, without (left) and with (right) the global optimal estimator ($H = 5.44\,bpp$).

Figure 7. Result for the Identity transform, without (left) and with (right) the global optimal estimator ($H = 7.26\,bpp$).

5. CONCLUSION

As the image examples clearly show the yielded gain by using the derived estimation formula is not only of theoretical matter. Using the fine-detunable CCT and the globally optimal estimation we yield a combined source-channel coding system which permits fine-tuned trade-offs between error robustness and bit rate increase in a manner which is transparent to the following stages (e.g. quantization and entropy coding) of the overall communication system.

The experiments using the identity transform shows that the derived estimation formula may be useful in other application, e.g. replacing bad pixels in an image.

Figure 8. Result for the CCT transform ($\varepsilon = 0.0$), without (left) and with (right) the global optimal estimator ($H = 5.26\, bpp$).

Figure 9. Result for the CCT transform ($\varepsilon = 0.02$), without (left) and with (right) the global optimal estimator ($H = 5.35\, bpp$).

REFERENCES

1. V. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," pp. 674–678.
2. V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of gaussian vectors," in *Data Compression Conference*, pp. 388–397, 1998.
3. Y. Wang, M. Orchard, and A. Reibman, "Multiple description coding using pairwise correlating transforms: analysis for gaussian sources and application to images," 1999.

Figure 10. Result for the CCT transform ($\varepsilon = 0.08$), without (left) and with (right) the global optimal estimator ($H = 5.81\, bpp$).

4. J. K. R. Arean and V. Goyal, "Multiple description perceptual audio coding with correlating transforms," 2000.
5. V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine* **18**, pp. 74–93, September 2001.
6. N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice Hall Professional Technical Reference, 1990.
7. T. Sikora and H. Li, "Optimal block-overlapping synthesis transforms for coding images and video at very low bit-rates," *IEEE Transactions on Circuits and Systems for Video Technology* **6**(2), pp. 157–167, 1996. shelf redesigns the inverse DCT for the case where only a few coefficients are transmited.