<u>Subband coring for image noise reduction.</u>

Edward H. Adelson

Internal Report, RCA David Sarnoff Research Center, Nov. 26 1986.

Let an image consisting of the array of pixels, E(x,y), be denoted **E** (the boldface indicating that it is a discrete function over (x,y)). Likewise, let a kernel with taps A(x,y) be denoted **A.** The image can be decomposed into a set of N filtered subimages, $\mathbf{S}_i$, by convolution with the set of N kernels, $\mathbf{A}_i$. Let * indicate convolution; then the decomposition is:

$$\mathbf{E} = \mathbf{S}_0 + \ldots + \mathbf{S}_{N-1}$$

where

$$\mathbf{S}_i = \mathbf{E}^*\mathbf{A}_i$$

We do not assume that the subimages have been decimated. Thus each subimage has the same number of pixels as the original image, and the total number of pixels in the decomposition is N times the number in the original image.

The kernels would normally be chosen to produce a decomposition into a "useful" set of subimages. In the case of noise coring, this will usually mean that the kernels are selective for orientation and scale, i.e. that they select out limited patches in the spatial frequency domain. For example, we might choose $\mathbf{A}_1$ to select for vertical energy, $\mathbf{A}_2$ to select for horizontal energy, and $\mathbf{A}_3$ to select for diagonal energy, as shown by the spectra in figure 1(a). $\mathbf{A}_0$ contains the remaining lowpassed energy. The decomposition is shown as a flow diagram in figure 1(b).

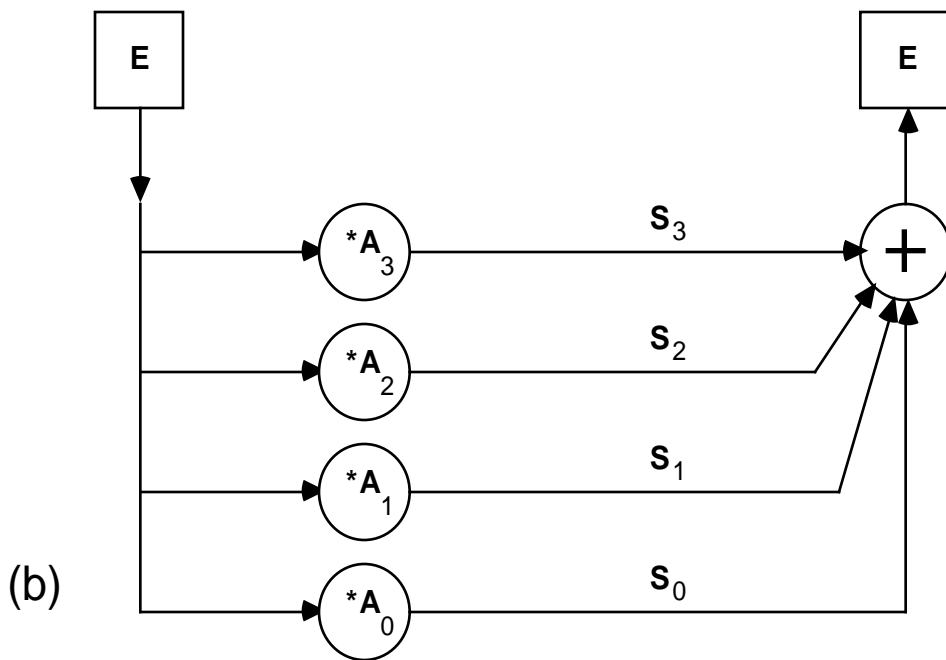| $\mathbf{A}_0$ | $\mathbf{A}_1$ | $\mathbf{A}_2$ | $\mathbf{A}_3$ |
| Lowpass | Vertical | Horizontal | Diagonal |

(a)

(b)
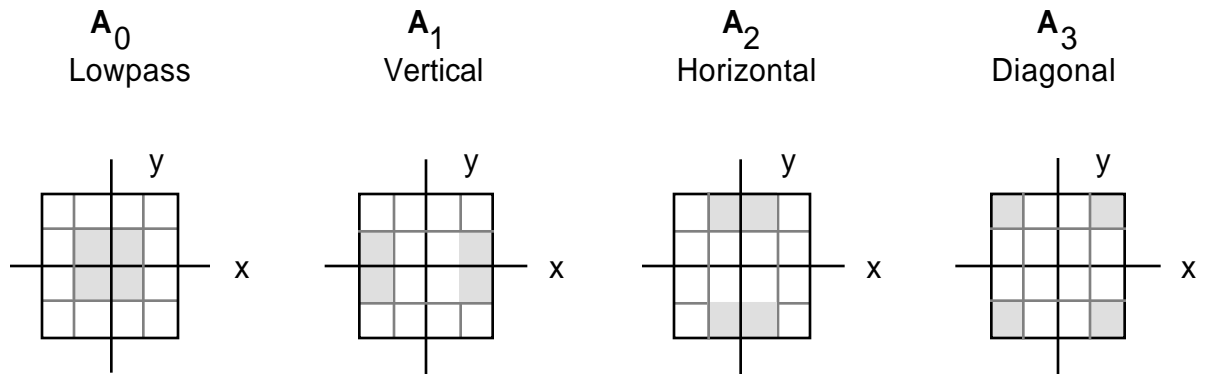
Figure 1: (a) the spatial-frequency spectra for a decomposition into lowpass, vertical, horizontal, and diagonal energy. (b) A flow diagram of the decomposition and reconstruction. The original image, $\mathbf{E}$, is decomposed into the 4 subimages, $\mathbf{S}_0,...\mathbf{S}_3$, by by filtering with the kernels $\mathbf{A}_0,...\mathbf{A}_3$. These subimages sum to reconstruct the original, $\mathbf{E}$.

(Observe that it is sufficient, but not necessary, that the kernels themselves sum to the unit impulse , $\delta(x,y)$).

Suppose now that each kernel A_i can be expressed as the convolution of a kernel B_i with itself:

$$A_i = B_i * B_i.$$

Then one can define a symmetrical subband transform as follows:

<u>Forward transform:</u>
Let
$$T_i = E * B_i.$$

Then the transform is:
$$E \implies \{ T_0, \ldots, T_{N-1} \}$$
Thus transform consists of the set of subband images, T_i.

<u>Inverse transform:</u>
To reconstruct the image from the transform, apply the kernels B_i once again to the corresponding images T_i and sum the resulting images S_i.

$$T_0 * B_0 + \ldots + T_{N-1} * B_{N-1} = S_0 + \ldots + S_{N-1}$$
$$= E$$

The transform is illustrated in figure 2. The image E is filtered through a bank of parallel filters, indicated by the convolving kernels B_i. The resulting filtered images are refiltered by the same kernels, and then summed to reconstruct the original image.
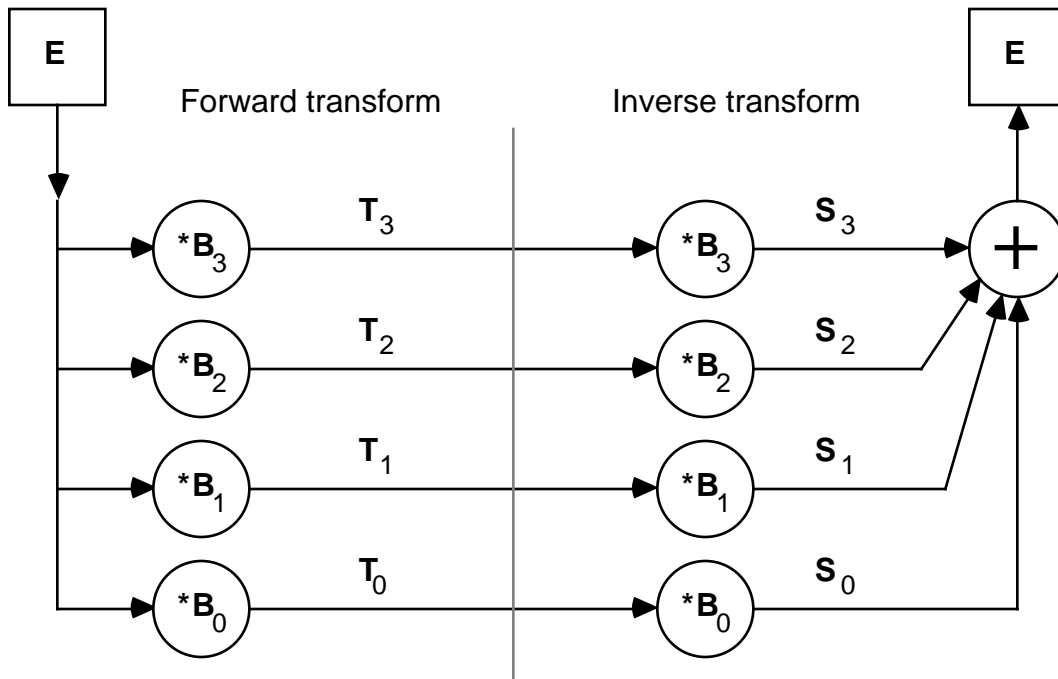
Figure 2: A symmetrical subband transform. Image **E** is filtered by convolution with kernels **B**$_0$,...**B**$_3$, to form the transform images **T**$_0$...**T**$_3$. A second filtering leads to a set of subimages **S**$_0$...**S**$_3$ that can be summed to reconstruct the original image, **E**.

In a typical subband transform, one of the filters is lowpass, containing all of the DC component, while the others are bandpass, with no DC. We will assume that the kernel **B**$_0$ is lowpass, as are the associated images **T**$_0$ and **S**$_0$.

The basis functions of the subband transform are the kernels themselves, taking centers at all positions in the image. That is, the transform can be considered to decompose the image into a sum of subimages, where the subimages themselves are sums of kernels repeated along the (x,y) grid:

$$\mathbf{E} = \sum_i \mathbf{S}_i$$

where

$$S_i(x,y) = \sum_{u,v} T_i(u,v)B_i(x-u,y-v)]$$

If the basis functions are selective for useful image information such as edges or lines, then the transform can be used to reduce visible noise by applying a static non-linear coring function. A typical coring function is shown in figure 3. It is assumed that the values of the coefficients to which it is applied cover the arbitrary range (-127 to 127). If a coefficient is near zero, its value is likely to represent noise and therefore it is attenuated. If a value is large, it is likely to represent legitimate image information and therefore it is left unchanged or slightly amplified. There is a smooth transition between the "coring" region and the "peaking" region. The precise shape of the best coring function will depend on the characteristics of the signal and the noise, but the coring function will have this general form.
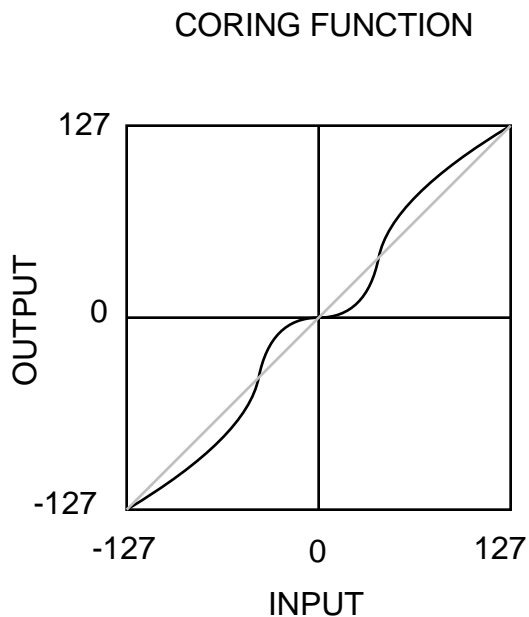
CORING FUNCTION

Figure 3: A typical coring function, which acts as a static non-linearity on each pixel value of a bandpass transform image. Values near zero are attenuated, while values far from zero are left unchanged or are amplified.

Coring can be usefully applied to all of the bandpass images in the transform, i.e. all of the images except for the lowpass image T_0. The process

is shown in figure 4. Denote the coring functions $c_i(.)$, the cored transform images $\tilde{T}_i$, and the resulting cored subimages $\tilde{S}_i$. That is,

$$\tilde{T}_i\ (x,y) = c_i\ (\ T_i\ (x,y)\ )$$

and

$$\tilde{S}_i = \tilde{T}_i * B$$

The final cored image is then

$$\tilde{E} = \sum_i [\tilde{S}_i]$$
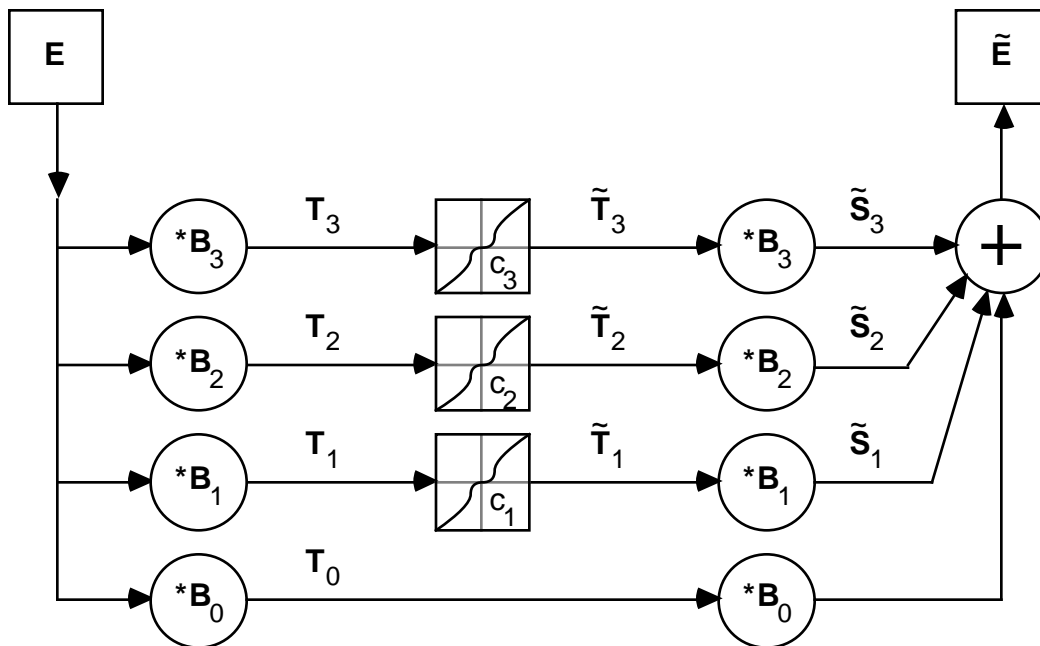
Figure 4: Flow diagram for coring an image. The original, **E**, is transformed to the images $T_0,...T_3$. The three bandpass images, $T_1,...T_3$, are then cored with the static non-linearities, $c_1, ..., c_3$. Reconstruction proceeds normally. The final cored image, $\tilde{E}$, should have less noise.

Coring is more effective when carried out at multiple spatial frequency bands. The previous description applies to coring at a single spatial frequency range determined by the kernels B_i. The lowpass transform image,

T_0, was not cored, and so all of its information (both signal and noise) is passed unchanged. This information can be further subdivided by applying a second subband transform with kernels tuned to lower frequencies, and coring can be applied to the resulting transformed images T'_i.

The most efficient way to achieve this is hierarchically, as shown in figure 5. At the first level, the image **E** is transformed into the images T_0, T_1, T_2, and T_3. Since T_0 is lowpass, it cannot be directly cored. However it can be used as input to a second coring stage, where all the kernels operate at lower frequencies. The second stage kernels are denoted B'_i, the transform images T'_i , the coring function c'_i , the cored transform images $\tilde{T}'_i$, and the cored subimages $\tilde{S}'_i$. The cored version of T_0 , denoted $\tilde{T}_0$, is then combined with the cored transform images from the first stage, and the final cored image is reconstructed.

This process can be repeated for several stages; each stage will remove noise in a lower frequency band than the previous stage. In practice there is usually little advantage to repeating the process at more than 3 frequency bands.
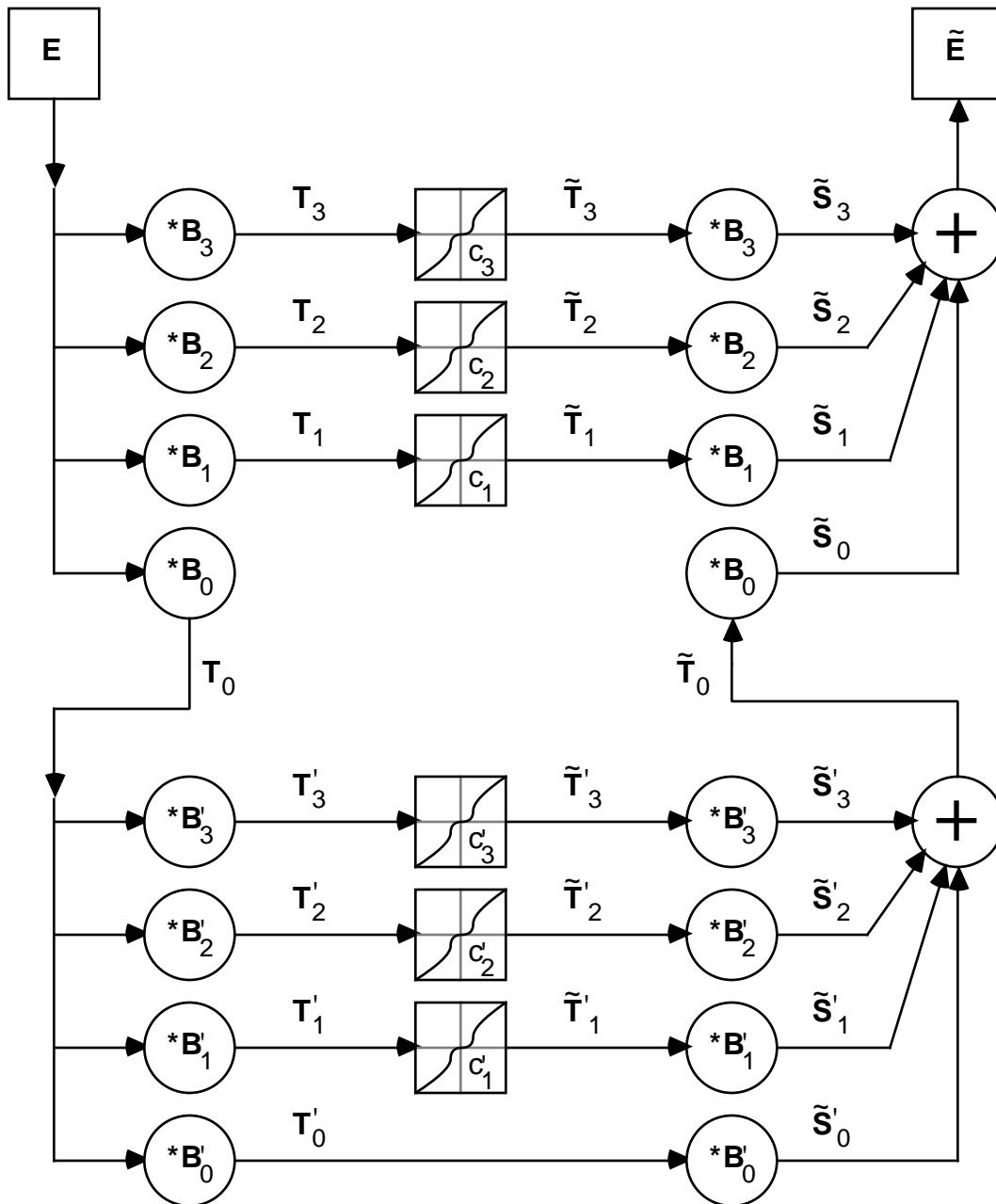
E

$\tilde{\mathbf{E}}$

$\mathbf{T}_3$ $\tilde{\mathbf{T}}_3$ $\tilde{\mathbf{S}}_3$

$*\mathbf{B}_3$ c_3 $*\mathbf{B}_3$ $+$

$\mathbf{T}_2$ $\tilde{\mathbf{T}}_2$ $\tilde{\mathbf{S}}_2$

$*\mathbf{B}_2$ c_2 $*\mathbf{B}_2$

$\mathbf{T}_1$ $\tilde{\mathbf{T}}_1$ $\tilde{\mathbf{S}}_1$

$*\mathbf{B}_1$ c_1 $*\mathbf{B}_1$

$*\mathbf{B}_0$ $*\mathbf{B}_0$ $\tilde{\mathbf{S}}_0$

$\mathbf{T}_0$ $\tilde{\mathbf{T}}_0$

$\mathbf{T}'_3$ $\tilde{\mathbf{T}}'_3$ $\tilde{\mathbf{S}}'_3$

$*\mathbf{B}'_3$ c'_3 $*\mathbf{B}'_3$ $+$

$\mathbf{T}'_2$ $\tilde{\mathbf{T}}'_2$ $\tilde{\mathbf{S}}'_2$

$*\mathbf{B}'_2$ c'_2 $*\mathbf{B}'_2$

$\mathbf{T}'_1$ $\tilde{\mathbf{T}}'_1$ $\tilde{\mathbf{S}}'_1$

$*\mathbf{B}'_1$ c'_1 $*\mathbf{B}'_1$

$\mathbf{T}'_0$ $\tilde{\mathbf{S}}'_0$

$*\mathbf{B}'_0$ $*\mathbf{B}'_0$

Figure 5: Two stage coring. The low-pass transform image $\mathbf{T}_0$ undergoes a second stage of coring by a set of kernels $\mathbf{B}'_0, ... \mathbf{B}'_3$. These kernels are tuned to lower spatial frequency than the first set, and so core out noise in a lower frequency band.

We will now describe a simple concrete example of such a two-level coring system, using 2x2 Hadamard functions as the kernels. These are not

very good kernels because their poor filter selectivity; but they offer an example that is easily understood. The first-level kernels are as follows:

$$\mathbf{B}_0 = 1/4 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \mathbf{B}_1 = 1/4 \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{B}_2 = 1/4 \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \qquad \mathbf{B}_3 = 1/4 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

And for the second level, the same kernels are simply padded with zeroes to spread their taps to double the distance, thereby reducing their frequency tuning by one octave:

$$\mathbf{B'}_0 = 1/4 \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \mathbf{B'}_1 = 1/4 \begin{bmatrix} -1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{B'}_2 = 1/4 \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \qquad \mathbf{B'}_3 = 1/4 \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Note that at the second level the kernels are being applied to an image that has already been convolved with the lowpass kernel $\mathbf{B}_0$. Thus the effective kernels at the second stage are actually given by the convolutions of the new kernels, $\mathbf{B'}_i$, with $\mathbf{B}_0$:

$$\mathbf{B}_0 * \mathbf{B}_0' = \; 1/16 \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad \mathbf{B}_0 * \mathbf{B}_1' = \; 1/16 \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{B}_0 * \mathbf{B}_2' = \; 1/16 \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} \qquad \mathbf{B}_0 * \mathbf{B}_3' = \; 1/16 \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$$

The same process of padding and convolution can be repeated at further stages. Large effective 2-D kernels are thus generated, using the principles of hierarchical discrete correlation previously outlined by Burt (1981).

More effective coring requires kernels with better tuning. A good set are the quadrature mirror filters (QMF's) that have been widely used in speech coding (Crochiere and Rabiner, 1983). A pair of high-pass and low-pass QMF kernels can be separably combined to form a set of four 2-D kernels (cf. Woods and O'Neil, 1986). Note that QMF kernels are designed to be used with decimation, but that we do not actually decimate in our coring procedure. This is because we find that decimation introduces "jaggies" and other aliasing artifacts when combined with the non-linear operation of coring. Since we do not decimate we are effectively working at double the necessary linear sample density at the first stage, and at quadruple the necessary linear sample density at the second stage. This extra density leads to extra computations, but prevents aliasing artifacts.

Here is an example of a 12-tap low-pass and high-pass QMF pair, taken from Johnson (1980):

Lowpass: Highpass:

-.003809   .003809
.018857   -.018857

-.002710	.002710
-.084696	.084696
.088470	-.088470
.484389	-.484389
.484389	.484389
.088470	.088470
-.084696	-.084696
-.002710	-.002710
.018857	.018857
-.003809	-.003809

The QMF kernels that have been published have an even number of taps, and the high-pass kernels are odd-symmetric. We have also derived kernels that are even symmetric, with an odd number of taps; these kernels are different from those that have been published, and tend to be more compact. These filters were derived numerically to satisfy the three constraints: (1) narrow spatial frequency tuning, (2) spatial compactness, and (3) spectral completeness (i.e. ability to reconstruct the original image accurately from the subimages).

A 5-tap kernel pair is shown here:

Lowpass:	Highpass:
-.0516	-.0516
.2500	-.2500
.6032	.6032
.2500	-.2500
-.0516	-.0516

A 7-tap kernel pair is shown here:

Lowpass:	Highpass:
-.0052	.0052
-.0516	-.0516
.2552	-.2552

.6035	.6035
.2552	-.2552
-.0516	-.0516
-.0052	.0052

By combining these two 1-D kernels separably, in all four pairings, one can create four 2-D kernels, selective for low-pass, vertical, horizontal, and diagonal energy. Because these kernels are spatially compact, they lead to coring that is well localized, without ringing artifacts. And because the kernels have compact spectra, they are good at selecting out oriented image information within each frequency band.

Asymmetric transforms.

We note that it is also possible to use asymmetric transforms, where the sampling functions are different from the basis functions. We previously assumed that the original set of kernels, A_i, could be expressed as the convolution of B_i with itself. This amounts to saying that the sampling functions are the same as their corresponding basis functions (up to a scaling), a fact that will hold for orthogonal basis sets. More generally we can assume that the A_i are expressible as the convolutions of sampling/basis pairs that need not be equal. This occurs when the sampling and basis functions are related by the pseudo-inverse.

The asymmetric transform is illustrated in figure 6. Let the sampling kernels be b_i and the basis kernels be B_i . Then we let

$$T_i = E * b_i.$$

and the forward transform is:

$$E \implies \{ T_0, \ldots, T_{N-1} \}$$

For the inverse transform, the kernels B_i are applied to the corresponding images T_i to give the subband images S_i. These are summed to reconstruct the original image. That is,

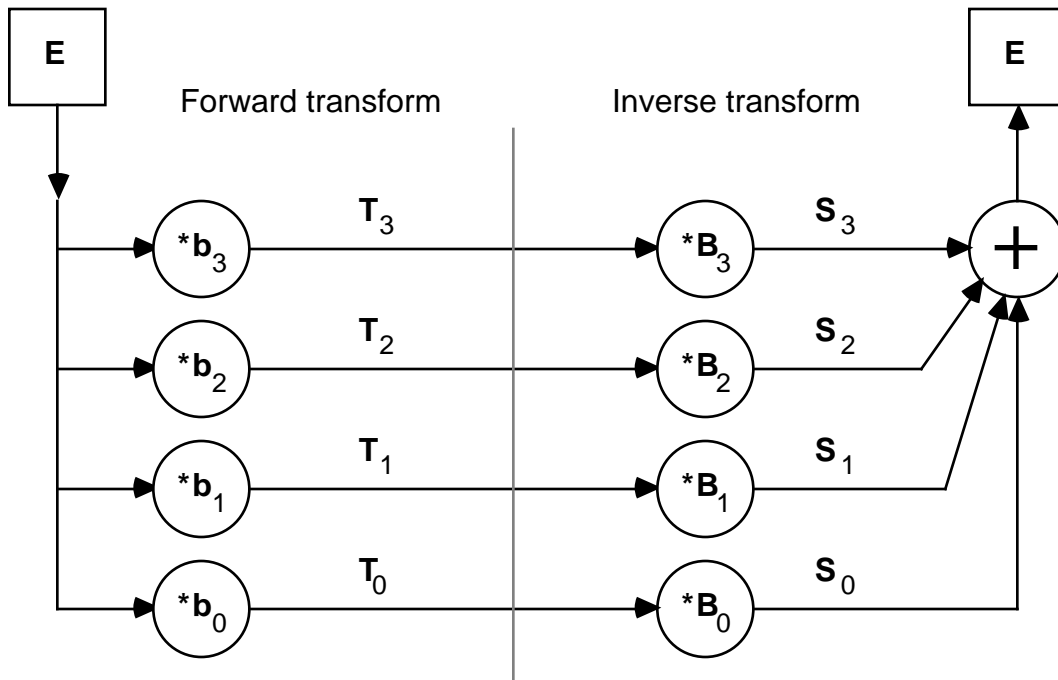$$T_0 {}^*B_0 + \ldots + T_{N-1} {}^* B_{N-1} \;=\; S_0 + \ldots + S_{N-1}$$
$$= \; E$$

Figure 6: An asymmetric subband transform. The transforming kernels are $\mathbf{b}_0, \ldots \mathbf{b}_3$, while the kernels used for reconstruction are $\mathbf{B}_0, \ldots \mathbf{B}_3$.

References:

Burt, P J "Fast filter transforms for image processing," *Comput Gr. Image Proc., 16*, 20-51(1981).

Crochiere, R E, and Rabiner, L R, *Multirate digital signal processing*, Englewood Cliffs: Prentice-Hall (1983).

Johnston, J D "A filter family designed for use in quadrature mirror filter banks," *Proc. 1980 IEEE Int. Conf. Acoust. Speech Signal Process*, pp 291-294. (1980).

Woods, J W, and O'Neil, S D, "Subband coding of images," *IEEE Trans. Acoustics, Speech and Sig. Proc., ASSP-34*, 1278-1288 (1986).