

DATA COMPRESSION AND WAVELET TRANSFORMS

G.M. RICHTER¹, M. CAPACCIOLI², G. LONGO² and H. LORENZ¹

¹ *Astrophysikalisches Institut Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany*

² *Osservatorio Astronomico Capodimonte, via Moiariello 16, I-80131 Napoli, Italy*

ABSTRACT.

Efficient data compression needs analyzing functions to recognise the local resolution of the signal. They are provided by the wavelet concept. The optimal wavelet (best information concentration) is defined by the image model of the application. For the most common images in astronomy the H-transform is optimal (in the sense of Karhunen-Loeve transform). The role of the H-transform in 2-dimensional processing is the same as the Haar-transform in 1-dimension, but it is not the '2-dimensional Haar-transform' found in text books.

1. Data Compression Requirements

There are two principal types of data compression: deterministic and statistical compression.

The deterministic compression, which allows the exact recovery of the original data, is essentially a matter of coding theory and will not be dealt with here. For efficiently compressing astronomical images, we need statistical methods; in this paper we will try to show why and how to do this.

The telescopes planned for the future provide extreme amounts of data (e.g. the LITE provides about 10 GBytes per night), but on the other hand these data have a very high redundancy, $r = 0.99 - 0.999$ (for details see Richter 1978). This means that the mean information content of a pixel is very low; in other words, if one knows the intensity of one pixel, the neighbours can be forecast with very high probability because almost every pixel is a background pixel and its neighbours are background, too. This means the pixels are strongly correlated. Thus, the three terms correlation, forecastability, and diluted information are only different aspects of the same phenomenon.

However, the information cannot be concentrated by simply keeping only 1 out of 100 or more pixels, because of the following two reasons:

- 1) Because of the noise, a single pixel of the background represents the background only with a considerable error. What we need is a mean value over a large number of pixels. On the other hand, this is already all the information we need, and there is no need to recover the exact value of every pixel (statistical compression!);

- 2) the signal process in astronomical images is strongly non-stationary, i.e. besides the low information background there are areas (e.g. stars, galaxies) where information is concentrated on a short scale length or (e.g. in the outer parts of galaxies) with intermediate resolution.

Hence, a data compression algorithm has to recognize the local signal scale lengths in order to separate signal from noise optimally. Such an algorithm, based on a transform, was developed and applied many years ago (Fritze et al 1977; Richter 1978). This transform is of the type called 'wavelet transform', which has received increasing attention for some years.

2. Wavelet Transforms

The wavelet transforms belong to the group of local transforms, i.e. they are localised in space and in frequency domain. This means that a particular transform coefficient is defined by the image points of a limited area only, in contrast e.g. to the Fourier transform which is a global transform where every coefficient is affected by every image point. On the other hand, a wavelet coefficient does not represent a sharp frequency in the frequency domain (like a Fourier coefficient), but in an extended frequency band.

The general form of a wavelet transform can be written (for simplicity in one dimension) as;

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x)g\left(\frac{x-b}{a}\right) dx \quad (1)$$

where $f(x)$ is the input function (image) and g is the 'analyzing wavelet'. Thus, every coefficient $C(a,b)$ is produced by the analyzing wavelet which is shifted by b and scaled by a .

In order to specify the transform for practical use, a and b must be chosen. The parameter a is usually chosen as

$$a = 2^i, \quad i = 0, 1, \dots \quad (2)$$

With the increment

$$\Delta b = 2a \quad (3)$$

a complete, orthogonal transform can be constructed with a typical hierarchical structure; one half of the coefficients have the highest resolution, one quarter have half this resolution, one eighth have half the resolution again, etc. Altogether the same number of coefficients as image points appear.

In principle, there is a many times indefinite manifold of analyzing wavelets and, thus, of wavelet transforms. Figure 1 shows two of them; on the left side a wavelet which can be found in more detail by Mallat (1989), and on the right side the Haar-wavelet (Haar 1910). In both cases the upper panel is the 'point spread function' in the space domain, and the lower panel is the 'transfer function' in the frequency domain.

The Mallat-wavelet is very well localised in frequency domain but less in space domain. The contrary is true for the Haar-wavelet. Intuitively one would expect that the Mallat-wavelet would be more suitable to analyze signals with semiperiodic characteristics, whereas the Haar-wavelet adapts better for signals preferably defined in space (e.g. stars). This raises the question of whether one can find an optimal wavelet.

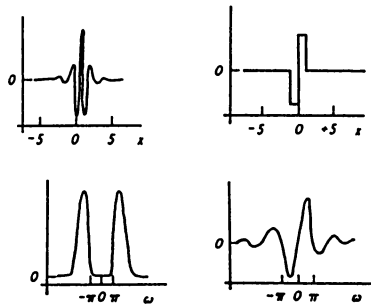


Figure 1. Analysing wavelets of Mallat (left side) and Haar (right side). In both cases, upper panel shows the impulse response in the space domain, lower panel the transfer function in the frequency domain.

Obviously, this depends on the particular application. Thus, we have to define:

- 1) a criterion of optimality; and
- 2) criteria which describe adequately the expected signal.

For 1) we can say that a transform is optimal if it concentrates the information into the minimal number of coefficients. The remaining coefficients are essentially zero; this means that they are dominated by noise. The next section gives a short sketch how to find this optimal transform, and also how to introduce the characteristics of the particular application.

3. Optimal Transform Coding

If an image is treated as a matrix, it can be transformed to a diagonal matrix by standard methods of matrix algebra (for details see e.g. Lorenz et al. 1993), where all matrix elements vanish except the diagonal elements λ_i . The transform is established by the eigen-vectors of this particular image, and is called an eigenvalue transform.

The eigenvalue transform has the following important property: if the i are assigned to the λ_i so that the λ_i form a decreasing function of i , then this function is the lower boundary of the respective functions of all possible transforms. This is due to the fact that all λ_i are mutually uncorrelated as a consequence of the diagonalisation. And this again causes each λ_i to have the largest possible information content. Thus, with the eigenvalue transform we have found the transform which concentrates the information optimal. Unfortunately, it is useless for data compression, because it has different eigenvectors for every image and, with respect to the present paper, it is not of the wavelet type.

This problem may be circumvented if we do not calculate the eigenvalues and eigenvectors for each individual image but instead estimate the averaged or expected eigenvectors of a whole

class of images, and then transform the images in the conventional way with this set of eigenvectors as transform matrix. Then the result is not exactly a diagonal matrix. However, all the elements outside the main diagonal are negligibly small; and the elements of the diagonal are not exactly decorrelated but are almost so. This approach is called the Karhunen-Loeve transform.

Obviously, this transform is only as good as the estimate of the expected eigenvectors. Usually they are estimated from a statistical model of the images, e.g. a Gauss-Markov process. However, there may be doubts whether these are realistic models of real images. We will demonstrate here another way to define an image model and introduce it into the Karhunen-Loeve transform.

Instead of transforming a whole image, we can divide it in blocks and transform the blocks separately. In order to begin with the simplest possible case, we use blocks of two pixels only (which are neighbours in x) and call the pixel values in every block γ_1 and γ_2 .

Now we introduce the first statement of our image model:

1. *neighbouring pixels are correlated.*

This means, if we draw every block as a point in a γ_1/γ_2 diagram (Fig. 2), then blocks are strongly concentrated along the main diagonal. Only blocks touching stars etc. provide points a bit farther from the diagonal.

Now we decorrelate our data by rotating the coordinate system $\gamma_1\gamma_2$ to $\alpha_1\alpha_2$ (i.e. by 45°) and get the transform:

$$\alpha_1 = \frac{1}{\sqrt{2}}(\gamma_1 + \gamma_2) \text{ and } \alpha_2 = \frac{1}{\sqrt{2}}(\gamma_1 - \gamma_2) \quad (4)$$

Note that α_1 and α_2 are now decorrelated and the variance of α_1 is much larger than that of α_2 .

However, our image model is not yet complete and nor is our Karhunen-Loeve transform. The next statement of the model is:

2. *The correlation is isotropic in x and y .*

Therefore, we take now quadratic blocks of 2×2 pixels in the configuration:

$$\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}$$

where $i = 1 - 4$ are the indices of γ_i . Then a 4-dimensional system with again only the main diagonal populated is the counterpart to diagram of Fig. 2.

The decorrelating transform is:

$$\begin{aligned} h_o &= \frac{1}{2}(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4) \\ h_x &= \frac{1}{2}(\gamma_1 - \gamma_2 + \gamma_3 - \gamma_4) \\ h_y &= \frac{1}{2}(\gamma_1 + \gamma_2 - \gamma_3 - \gamma_4) \\ h_c &= \frac{1}{2}(\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4) \end{aligned} \quad (5)$$

Instead of α_i we use here already a notation which takes into account the physical interpretation of the transform coefficients: h_x and h_y are obviously the gradient components in x and y direction, and h_c is a curvature term.

The last statement of our model is:

3. *The correlation extends to larger pixel distances with decreasing magnitude.*

Continuing as above we would get now very high dimensional γ_i -spaces. However, we realize that the 2×2 pixel blocks are already decorrelated inside by Equ. 5. Therefore, it is enough to

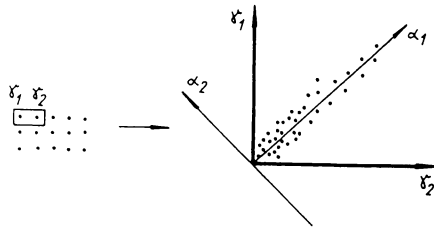


Figure 2. Typical co-occurrence diagram of neighbouring pixels in an image.

treat Equ. 5 as first order of the transform and then put the mean values h_o of 4 neighbouring 2×2 blocks into the Equ. 5 instead of the γ_i . This can be repeated order by order and stopped when neighbouring blocks are no longer correlated.

With this procedure we have established a complete orthogonal transform which we call H-transform. The H is chosen in honour of Haar (1910). However, we must emphasize that it is not the 2-dimensional Haar transform found in textbooks.

The h_x, h_y, h_c coefficients show the typical properties of analyzing wavelets, localisation, shift and scaling; but in contrast to the 1-dimensional case, we now have three types of wavelets. Hence, with the H-transform we have established a wavelet transform, which is suboptimal (in the sense of Karhunen-Loeve transform) for signals which obey the 3 criteria of the model cited above.

References

- Fritze, K., Lange, M., Möstel, G., Oleak, G. and Richter, G., 1977. *Astron. Nachr.*, **298**, 189.
 Haar, A., 1910. *Math. Ann.*, **69**, 331.
 Lorenz, H., Richter, G., Capaccioli, M. and Longo, G., 1993. *Astron. Astrophys.*, **277**, 321.
 Mallat, S., 1989. *IEEE Trans. Pattern An. and Machine Intell.*, **11**, 174.
 Richter, G., 1978. *Astron. Nachr.*, **299**, 282.