Wavelet packet representation of textured regions

Examensarbete utfört i Bildkodning
vid Tekniska Högskolan i Linköping
av

Jens Larsen

Reg nr: LiTH-ISY-EX-3022

Handledare: **Anna Linderhed**
Robert Forchheimer

Examinator: **Robert Forchheimer**

Linköping - 2001-03-27

Avdelning, Institution
Division, department

Department of Electrical Engineering

Datum
Date
2001-03-27

Språk
Language

❑ Svenska/Swedish
■ Engelska/English

❑ _______________

Rapporttyp
Report: category

❑ Licentiatavhandling
■ Examensarbete
❑ C-uppsats
❑ D-uppsats
❑ Övrig rapport

URL för elektronisk version
http://www.icg.isy.liu.se/

ISBN

ISRN

Serietitel och serienummer
Title of series, numbering

ISSN

LiTH-ISY-EX-3022

Titel
Title

Wavelet packet representation of textured regions

Författare
Author

Jens Larsen

Sammanfattning
Abstract

This master thesis examines a way of separating textured regions in an image and representing them using wavelet packets. Highly textured images tend to have a spectral distribution with much energy content in the mid-bands. Wavelet packet representation is thereby a suitable basis for the image and an extended form of tiling the spatial-frequency domain is suggested along with operations to manipulate the representation. This over-complete form of representation is called stacks and the operations are guided by selection masks.

The method is demonstrated in experiments and the extension of non-linear operations is discussed. The literature study contains a commented list.

Nyckelord
Keywords

image coding, wavelet packet, texture, separation, stacks

Abstract

Wavelet packet representation of textured regions

This master thesis examines a way of separating textured regions in an image and representing them using wavelet packets. Highly textured images tend to have a spectral distribution with much energy content in the mid-bands. Wavelet packet representation is thereby a suitable basis for the image and an extended form of tiling the spatial-frequency domain is suggested along with operations to manipulate the representation. This over-complete form of representation is called stacks and the operations are guided by selection masks.

The method is demonstrated in experiments and the extension of non-linear operations is discussed. The literature study contains a commented list.

Table of Contents

1.	***Preface***	***Page 1***
2.	***Introduction***	***2***
3.	***Segmentation - Separation***	***4***
4.	***Theory of wavelet representation of images***	***6***
5.	***Experiment separating a highly textured image***	***17***
6.	***Experiment using selection masks***	***20***
7.	***Separation of a textured region***	***25***
8.	***Stacks, a suitable representation for separation***	***30***
9.	***Advanced selection masks***	***35***
10.	***Literature study***	***37***
A.	***Appendix***	***46***

1. Preface

Every line has to start somewhere. So has the guiding red line that marks the path of thought. This very spot is a good place to start the journey through this thesis and so you have. Welcome.

Separating an image into two additive images, based on the texture content in the image, is the goal of the master project summarised in this thesis. A new form of representation, the stack, is described and used in several examples. The stack is a generalised wavelet packet-decomposition and operations to manipulate the stack using selection masks are described.

Starting with the results gives you an overall view of what the destination will look like. It is followed by a guide to the wavelet representation of images and its historical development. The basic stack representation is a new form of representation designed to simplify the separation. Building on this representation, we select a few wavelet packets and separate one of Brodatz images, the brick wall. Advancing the concept further, the next step is to only select some of the coefficients from the decomposition. The image used to illustrate this is the well-known Lenna image.

Selecting wavelet packets and coefficients for transfer is the foundation of the concept presented in this thesis. Combining all of the previous methods the Barbara image is separated. It is the textured structure of her cloths, her left leg to be precise, that is separated. The result is good.

From there extensions to the separation scheme is presented. Non-binary selection masks and non-linear masks and their use in stacks are advanced topics that require understanding of all the previous chapters.

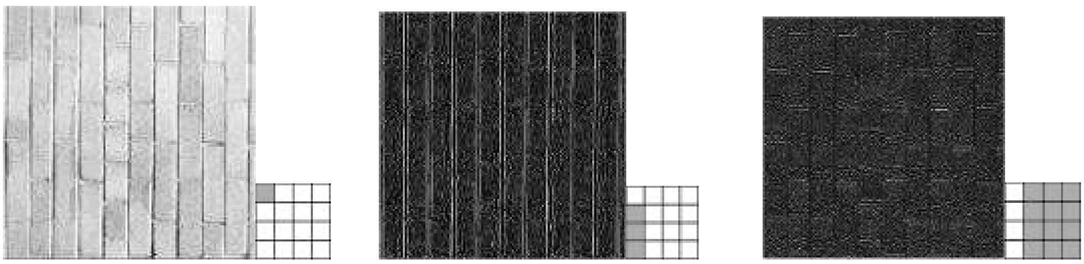
The extensive literature study is presented as a commented list. The appendix and chapter four on wavelet representation of images can be read independently and out of the context of the rest of this thesis.

Fasten your seatbelts.

Jens Larsen, larsen@isy.liu.se, LiTH, Sweden
Monday, 12 February 2001

2. Introduction

This thesis describes several experiments that each demonstrates principles and techniques used in the separation. Visually, all of them perform well and none of the experiments has been fine-tuned afterwards, but was carefully designed to capture the essence of the underlying idea in the first execution. This is a success. However, the small number of examples still leaves doubt about the generality of the techniques.

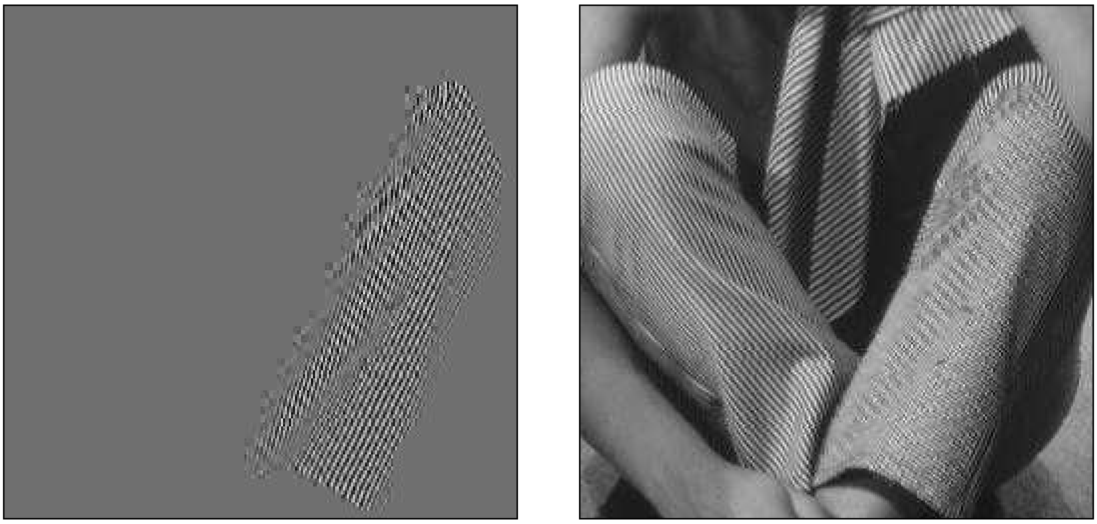


From chapter 5: Experiment separating a highly textured image. If these three images are added together, the original image is recovered.

The introduction of the representation form called *stacks* does not lack generality. It is probably the most general form of representation possible. The stacks are therefore classified into basic stacks and basis stacks to maintain the connection to conventional wavelet packet representation. If nothing else, the stack concept is the most fundamental issue of this thesis.

From chapter 6: Experiment using selection masks. The selection masks at several scales. Since they are binary, an AND-operation is all it takes to mask out the desired coefficients. The masks are independent of each other. Sharp edges are present at all scales, but other features are only present in some. This can be exploited in texture separation.

Operations within the stack have been developed and the link to texture separation is made using *selection masks* of different kinds. It is in this area that more work can be done to fully test this texture separation technique's potential. There are many questions to be answered about the different ways to create the masks, and also how the masks relate to each other over the scales. On top of that, in the later chapters there is a discussion of non-linear operations.



From chapter 7: Non-binary selection masks. Result from the separation. The striped leg to the right is separated from the initial image. To the right, the difference in detail and structure can be observed by comparing the texture on the right and left leg.

As usual, any attempt to answer a scientific question seems to generate even more questions. In this thesis, it is clearly demonstrated that it is possible to do the desired texture separation in a controlled way using stacks. In doing so, the generality of the method adds a higher degree of freedom in selection of parameters. This selection, in all its endless combinations, raises new questions.

I will end my introducing journey here and I hope that this thesis inspires others to follow some of the many lose ends.

3. Segmentation - Separation

Separation is a wider concept than segmentation

There is a difference between segmentation and separation. Separation is a wider concept than segmentation and to avoid confusion the term separation is used to describe the methods and results in this thesis.

Separation is an operation that divides an image into several parts. The individual parts are often subject to different coding methods or transmission requirements. This thesis deals with the issue of separating highly textured image content in a *spatial-frequency sense* and is not limited to disjunctive regions. The separated parts are additive, but the parts overlap spatially. This property disqualifies the method from being referred to as a segmentation scheme. The overlap is a natural consequence of working within the wavelet packet-representation.

Traditionally *segmentation* means a division of the image into smaller parts by simply cutting out certain areas. A more suitable name for these areas are *regions* since methods of this kind operate on the image itself with clearly defined borders, often rectangular. Many methods use the concept and refer to them as *region of interest*, ROI [5]. In the video coding standard MPEG4 a similar idea of video object plane, VOP, is used.

For texture-separation, wavelet packet-decomposition is promising. Most appealing is the multi-resolution property of the wavelet. The decomposition also forms a basis. Many other analysing functions do not form a basis and although they have supreme texture matching properties, control over the image is lost. For lossy compression, this might not be of importance, but unexpected results might be difficult to explain. A basis representation is a solid foundation from which deviations can be made. A new representation is presented and is called stacks.

Texture

Texture lacks a general definition. Somehow, an overall agreement on what is and what is not a textured area exists. Areas with certain texture can clearly be treated differently than non-textured areas. Imagine a portrait. The hair is a texture, but the face is not. It is possible to model the hair using orientation, thickness, curliness and so on. On the other hand, the region around the eyes is subject to demands on high detail. Scale is also important if a region is to be considered a texture or not. Texture representation using a model based synthesis can be successfully used [A13] but is troublesome to generalise. Classic texture analysis used textons as a fundamental building block for textures. The self-similarity of textured regions can be exploited using fractals.

Computer graphics use the term texture for the surfaces covering a tessellated 3D model. Sophisticated lighting models produce impressive result that looks very photo realistic. This definition is may be useful for model based image coding.

Texture is important to the human vision. It helps us to determine the size and shape of things. It also plays a big role in how sharp and detailed we perceive the image. Although textured areas can be very detailed, exact spatial localisation is not required by the human observer for recognition.

A certain degree of homogeneity in one or several properties is expected within a textured area. This can be in orientation, structure or pattern. The repetitiveness can be expressed in terms of frequency and this is found in the mid-bands often referred to as resonance-frequency. A wavelet packet representation that is optimised for these particular frequency bands can be constructed. A good example of this is the use of wavelet packets in the WSQ image-coding standard used for fingerprints in the criminal justice community. The image coding standard WSQ is described in the appendix.

4. Theory of wavelet representation of images

Development of wavelet based image coding in brief

State of the art image coders today are based on wavelet theory. The wavelet theory is a purely mathematical creation and its applications have been found very useful for image compression. Wavelet theory consists of a transform quite similar to the fourier transform or discrete cosine transform, DCT, and has nothing to do with image compression in itself. When the discrete wavelet transform, DWT, and its extensions like wavelet packets (flat decomposition instead of hierarchical), are used to derive the FIR filters actually used in practical applications the compression algorithm is referred to as being wavelet based. The filters can then be analysed in a more traditional way.

One of the fundamental differences between wavelet based coding and traditional subband coding, lies in the interpretation of the wavelet transform into classic subband terminology. Instead of discussing signal analysis in terms of filters and frequency bands, the wavelet transform uses a mathematical terminology such as wavelet functions, bases and subspace. Approximation space and detail space splitting is viewed as low pass and high pass filtering. The cascaded use of the very same filter on incrementally subsampled versions of the original image is one of the distinct marks of a wavelet based image coder. Another property is the multi-resolution analysis. From a principle point of view, this idea of several scales is a property that promotes the wavelet transform to something slightly more than just another transform.

History and development

J. Morlet [H1] was investigating a way to analyse seismological data and introduced the term *wavelet* in the early eighties. Wavelet means small wave and is a very suitable term for reasons that is obvious when we study the characteristic shape of a mother wavelet. The figure above is one example of a Daub4 wavelet, note that its mean is zero. The original term was *ondelettes*, which is French, because the pioneering work was made in France and Belgium. The reason for Morlet's studies was the need for an analysing tool for the very rapid changing frequency content of seismic data and fourier analysis was not sufficient. A few years later *Ingrid Daubechies* [H2] constructed families of orthonormal wavelet functions with compact support and *Stephane Mallat* [H3] put the wavelet transform in the framework of multi-resolution signal decomposition.

In 1993, *J.M Shapiro* [A15] published a compression algorithm, EZW, which outperformed all other compression algorithms. The algorithm very successfully tied together the necessary components of wavelet transform, selection of transform coefficients and entropy coding into a unit. Using Shapiro s genius idea of *irrelevant zerotrees* and refining it, the SPITH algorithm was developed by *Amir Said* and *William Pearlman* [A16]. SPITH is still one of the best compression algorithms in the research domain.

Future and industrial application of the JPEG2000

Today the industry has adopted the concept of wavelet theory. In the beginning of year 2000, the new image coding standard JPEG2000 was presented after several years of an international effort by the Joint Picture Expert Group. JPEG2000 outperforms the current JPEG-standard in all aspects; although the gap is not that great in terms of compression ratio, it has other properties such as error resilience and a loss-less mode. JPEG2000 is soon to be introduced in digital cameras and camcorders.

The current JPEG-standard is based on an entropy coded zig-zag scan of fixed 2D DCT-blocks. JPEG2000 has nothing to do with this method. Its generic form uses one of two wavelet filter banks and an advanced quantisation scheme called EBCOT, embedded block coder with optimised truncation.

Wavelets from a mathematical viewpoint

The transform is based around a selected square integrable function ψ(x). This function is called the *mother wavelet*. From this single function, translated and dilated wavelets can be derived. The wavelet functions may be either continuous or discrete. For our purposes, the discrete form is the only one of interest.

Wavelet functions

The translated (moved in space) and dilated (expanded in space) wavelet function derived from the mother wavelet are given by:

$$\psi_{mn}[k] = 2^{-\frac{m}{2}} \cdot \psi[2^{-m}k - n]$$

Equation 4.1

where n, m and k are integers, n is the translation factor and m the scaling factor.

A function that it is square integrable can then be written:

$$x[k] = \sum_{m} \sum_{n} \omega[m, n] \cdot \psi_{mn}[k]$$

Equation 4.2

where the inner products:

$$\omega[m,n] = \langle x, \psi_{mn} \rangle \equiv \sum_k x[k] \cdot \psi_{mn}[k]$$

Equation 4.3

are called the wavelet coefficients.

A requirement on ψ[k] is that it is square integrable, but equation (9.2) and (9.3) also imply that the functions generated in (9.1) are orthonormal. The same function is used in both the transform and the inverse. Two functions are orthonormal when their inner product is:

$$\langle \psi_{mn}, \psi_{m'n'} \rangle = \delta_{mm'} \cdot \delta_{nn'}$$

Equation 4.4

If the mother wavelet satisfy these equations, it forms an orthonormal basis into which we may transform our function and an inverse transform may uniquely be found. This is not different from any other transform. However, by adding two extra requirements to the mother wavelet ψ(x) the wavelet functions get properties that make the transform useful for practical applications like signal processing.

The mother wavelet should have *compact support*, which means that only a finite number of terms are non-zero. This is highly useful when calculating the sum in equation 9.3.

The second requirement is called the admissibility condition and implies:

$$\sum_k \psi[k] = 0$$

Equation 4.5

I.e. the wavelet functions mean value (0:th moment) is zero.

This simply means that the wavelet function is a zero mean function with compact support and hence exhibits some oscillatory behaviour. It is this fact that has given the function its name: One small wave, *ondelette* or in English, wavelet.

When functions that matched all these requirements were developed and the wavelet transform was put in the framework of multi-resolution analysis the importance of these wavelet bases became clear.

Multi-resolution analysis

Multi-resolution analysis is dealing with the representation of a function at several scales. Given the space $V_{-\infty}$ of all square integrable functions ($\in L_2$) approximations of this space can be produced recursively. From this follows that the approximation is a subspace to $V_{-\infty}$ and a signal x[k] in the subspace is an approximation of x. In this process, as we go to coarser levels, more and more information of the original signal is lost. The function $\phi[k]$, which generates the bases for these nested subspaces are called the *scaling function*, or sometimes, but rarely, the father wavelet. The scaling function must satisfy:

$$\sum_k \phi[k] = 1$$

Equation 4.6

i.e. the mean of the scaling function is one.

There exist many ways to perform this ladder of nested subspaces. Using a similar idea to the one found in the generation of wavelets from one mother wavelet (equation 4.1) is the signum of multi-resolution analysis. The use of a factor two in the dilation (expansion) is called dyadic and is especially useful in digital signal processing.

When going down to a coarser level (up in level of decomposition), from subspace V_m to V_{m+1} some information is lost to the complementary space W_{m+1}. Together the two subspaces V_{m+1} and W_{m+1} span exactly the same space as V_m. Just as $\phi_{mn}[k]$ forms a basis for V_m, there is a set of functions that form the basis for W_m. This basis is the wavelet basis earlier derived.

This pair of functions, one with zero mean and the other with unit mean, and their respective translations and dilations form a complete basis. The two sets of functions are tied to each other and given one of them, the other is uniquely determined. **This is a way to represent any square integrable function ($\in L_2$) at any dyadic level of resolution and the basis is given by the mother wavelet.**

Wavelets families

So, when Ingrid Daubechies constructed her family of wavelets that were both orthonormal and, more important, compactly supported and gave some guiding mathematical proofs a wide spectrum of wavelets was derived. Her filters are very common, mainly because they tend to have few coefficients and therefore are well localised in the spatial domain.

Other mother wavelet functions can be derived that satisfy some additional constrains. To represent smooth signals it seems to be a good idea to use smooth wavelets as well. A mathematical way of expressing this is in terms of moments:

$$\int x^l \phi(x) dx = 0, \qquad l = 1, \dots, L-1$$

$$\int x^l \psi(x) dx = 0, \qquad l = 1, \ldots, L-1$$

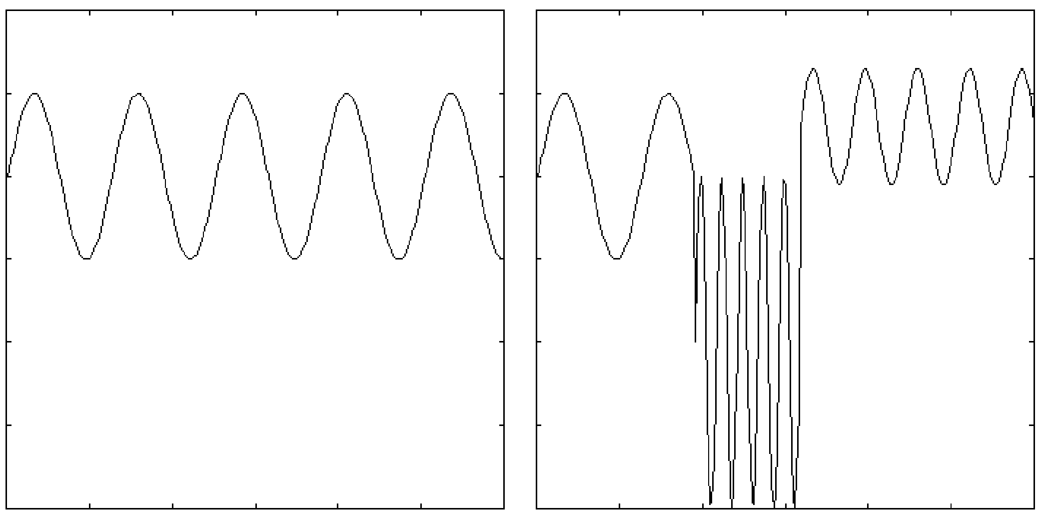
Note: l=0 would be the mean value, l=1 moment etc

Finding wavelets with several vanishing moments was solved by *R. Coifman* and these are called *Coiflets of order L.*

Other ways of constructing families of wavelets are by relaxing the requirement of compact support to only approximately compact support. A large set of functions that has exponential decay exists. Mostly used and quite easily derived are the *B-Spline wavelets.*

Wavelets from a signal processing viewpoint

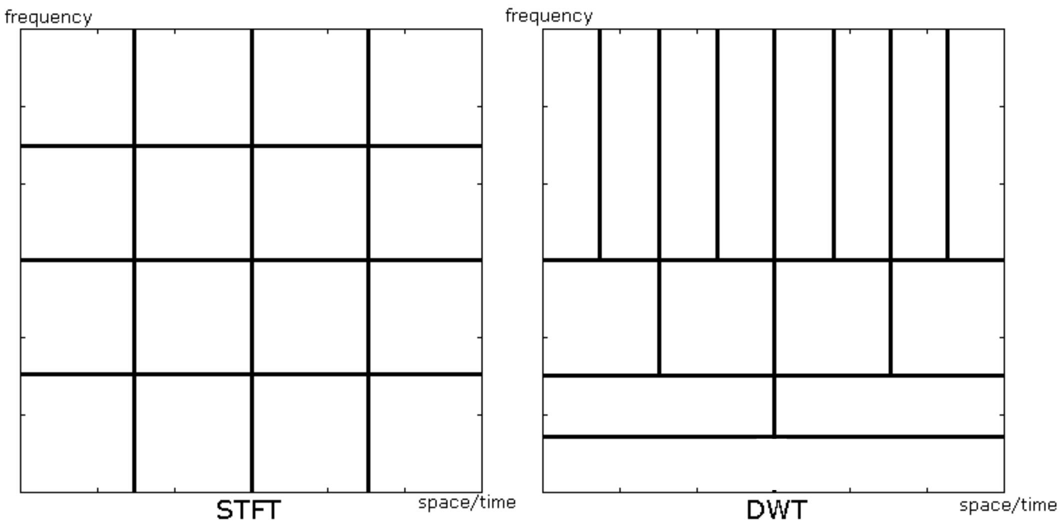
One of the basic ideas behind transform coding is that we can find basis functions, which fit well with the image to be coded. In this sense, we like to think that the image can be described in terms of frequencies. The fourier transform does a great job in finding the frequencies, but there is no special location associated with the frequency. It is similar to having a musical note score where it just says how many notes of A, B or G# there are in the piece, without the information about when and how long to play the notes.



On the left signal, fourier analysis is sufficient since it is periodic. The signal to the right needs another approach such as windowed fourier analysis or wavelet based methods.

Windowing of the analysing function

The most common work-around to this problem is to cut out only a piece of the infinite basis function. The way to cut is referred to as windowing the fourier transform and other names are short-time-fourier-transform, STFT or windowed fourier transform, WFT. The choice of window greatly affects the frequency response. Wavelets have this window built in and its size varies depending on the scale. This is very important since it is not possible to measure the frequency exactly and at the same time its location. The phenomena is a consequence of the *uncertainty relation* which is also found in quantum physics (Heisenberg's uncertainty relation). Equation 4.1 describes how the wavelet function is expanded in space. The figure below shows the tiling of the space-frequency plane and the area is held constant in the wavelet case, thereby trying to be as precise as possible in both domains. In other words, to measure high frequencies only a short filter is needed while low frequencies, slowly varying, require larger filter kernels and thereby affecting the precision in space or time.

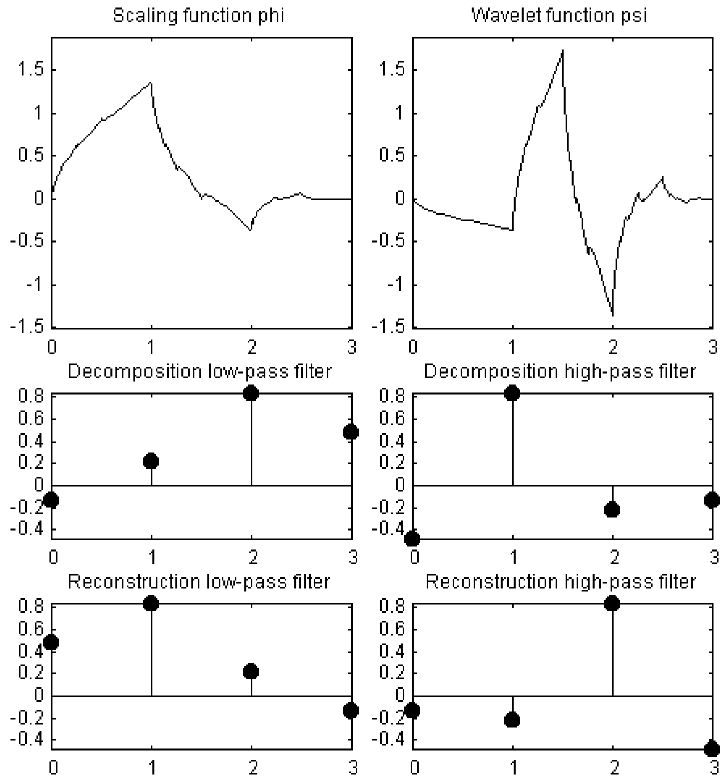


The tiling of the space - frequency plane for a one-dimensional signal differs between STFT and DWT (discrete wavelet transform). The window fixates the width in the short-time-fourier-transform, but in the wavelet transform case, the window varies with the frequency. The uncertainty relation sets the lower boundary of the tile size.

The use of a fixed windowing function is equivalent to cutting the image into fixed blocks before analysing the frequency. By decomposing the image using wavelets at different scales, block related artefacts are avoided. After decomposition into a approximation part (low-pass) and a detail part (high-pass) there is nothing stopping us from repeating the process once more. If we choose to only split the approximation part, again with the same filter the scheme is called *wavelet decomposition* and if we also split the detail part, the scheme is referred to as *wavelet packet decomposition.*

Implementation and filter banks

Practical implementations use filter banks constructed from the wavelet function. Since these functions have compact support, they can be used directly as FIR filters. The cascaded use of the same filter on the wavelet coefficients at different scales is one of the reasons that wavelet transforms are efficient to compute. An optimised algorithm can be used repeatedly. The combined cascaded filter is equal to a specific bandpass filter. This becomes clear if we consider how the wavelet functions are generated using equation 4.1.



Example of a compactly supported wavelet function and the associated FIR filters used to decompose and reconstruct one-dimensional signals. The wavelet function is the often-used Daubechies four tap filter.

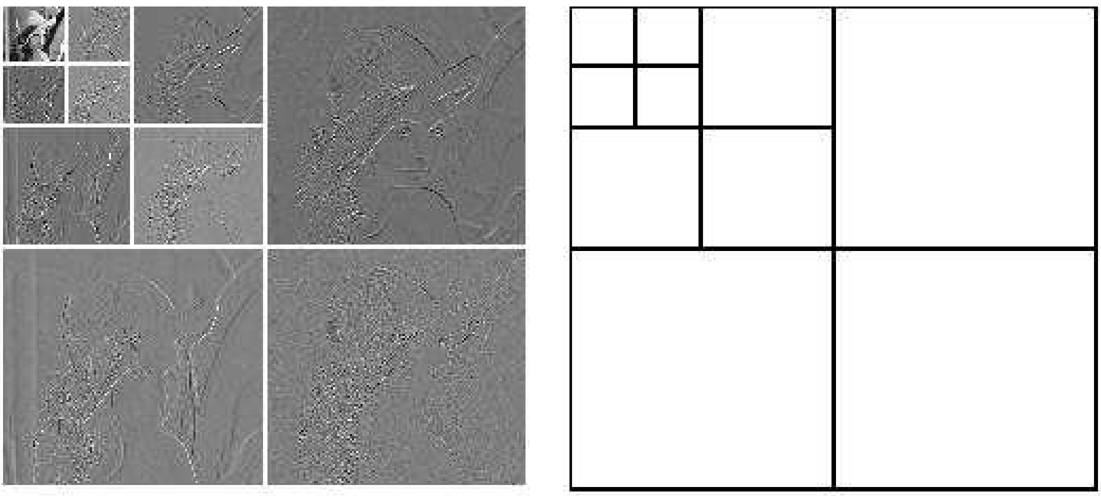
Some common wavelet FIR filter banks:

```
Haar (B-Spline):
» [lo,hi]=wfilters('haar')
lo =    0.7071    0.7071
hi =   -0.7071    0.7071

Daubechies:
» [lo,hi]=wfilters('db2');  % Comment N=2 is the number of vanishing moments
lo =   -0.1294    0.2241    0.8365    0.4830
hi =   -0.4830    0.8365   -0.2241   -0.1294

» [lo,hi]=wfilters('db3')
lo =    0.0352   -0.0854   -0.1350    0.4599    0.8069    0.3327
hi =   -0.3327    0.8069   -0.4599   -0.1350    0.0854    0.0352

Coiflets:
» [lo,hi]=wfilters('coif1')
lo =   -0.0157   -0.0727    0.3849    0.8526    0.3379   -0.0727
hi =    0.0727    0.3379   -0.8526    0.3849    0.0727   -0.0157

These FIR filter banks where generated using MatLab Wavelet toolbox.
```

Images are two-dimensional signals. Until now, the discussion has only dealt with one-dimensional functions and signals. A way of expanding these filters is to make them cartesian-separable in a straightforward construction. Given a low-pass filter h, and a corresponding high-pass filter g, four new filters can then be derived.

$$
\begin{aligned}
h_{LL}[k,l] &= h[k] \cdot h[l] \\
h_{LH}[k,l] &= h[k] \cdot g[l] \\
h_{HL}[k,l] &= g[k] \cdot h[l] \\
h_{HH}[k,l] &= g[k] \cdot g[l]
\end{aligned}
$$

The subscript denote the low-pass and the high-pass characteristics in x and y direction of the image.



Three levels of wavelet decomposition of the famous Lenna picture using the Haar filter. Only the low-pass band is split further. Coefficients have their magnitude coded in greyscale. On the right only the grid is shown, this way of representing the decomposition is often used.

The wavelet decomposition of Lenna in three levels shows the wavelet coefficients (not the reconstructed image) magnitude coded in greyscale. The upper left tile contains the low-pass band and clearly has the most energy, but on the other hand, it only contains 6.25 per cent of the original amount of data. If we store this low-pass band and only the significant wavelets coefficients in the other bands a data reduction is achieved. What is considered significant is one of the hard questions to answer in efficient image coding. Typically, only the coefficients with a magnitude above a certain threshold are regarded as significant. In an energy or PSNR sense, this is correct.

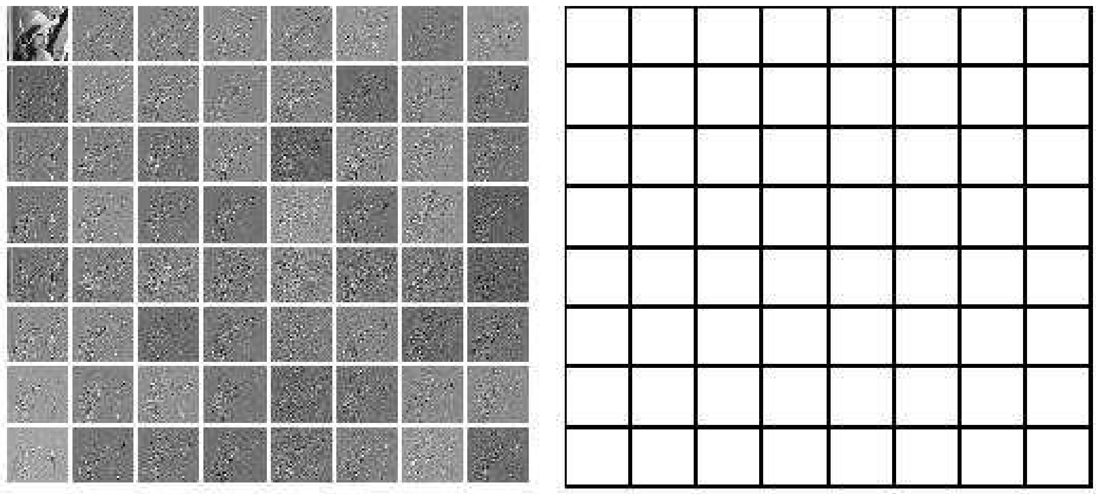
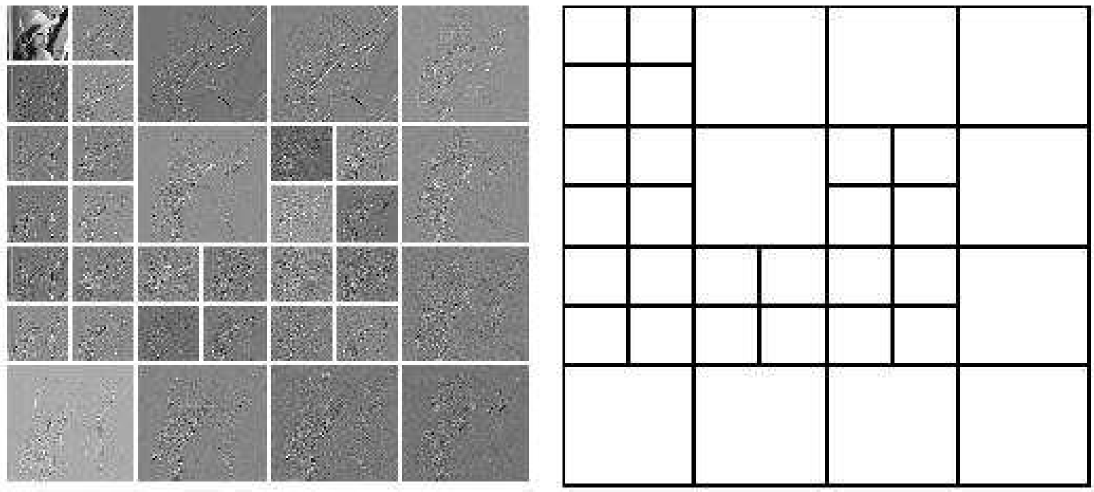


Figure shows complete wavelet packet decomposition in three cascaded levels. Note that the frequency order is not kept when cascading filters. Some literature refers to this total decomposition to be the only form of wavelet packet decomposition, disregarding the option of variations like the example below. This nullifies the benefits of multi-resolution since all the filter kernels have the same size.

This wavelet packet decomposition illustrates the difference in how the subbands are split. Whether it is beneficial to decompose the higher bands further or not depends on the image contents and highly textured images might have a resonance frequency that are in one of the middle bands. Federal Bureau of Investigations (FBI) stores fingerprints in a format that is a mixture of ordinary wavelet decomposition and wavelet packet decomposition. The compression scheme is called WSQ and it uses a 64-band decomposition before quantisation. A more detailed description of this scheme is found in the appendix.



There are many ways to split the image, this example is done using the best tree method using the Haar filter, and subband energy as cost function. The literature study indicates that this more general form of wavelet packet decomposition is referred to as best-tree splitting or alike. It is here the advantages of wavelet theory is used to the full, mixing different scales [A2], [A9].

Down sampling between the bands is done by simple decimation. Practically it is achieved by disregarding every second filter response, or more efficiently, by moving the filter two steps at a time before computing the convolution.

Bi-orthonormal wavelets

Ingrid Daubechies [H4] proved that if the filter banks are orthonormal then they could not be symmetrical. This is bad news from an image coding point of view, since the filter response should be independent of alignment of the image. It is possible to construct bi-orthonormal filters that are symmetrical, but they tend to have larger support, i.e. they have larger filter kernels. In practise, the non-symmetry is of less importance. Bi-orthonormal filters use a pair of filters where one is used for decomposition and the other for reconstruction.

The admissibility condition (the zero mean condition) gives at hand that the mother wavelet can not be a low-pass filter. On the other hand, it has to be orthogonal to the low-pass filter and this is calculated directly from the inner product (equation 9.4). If we construct a high-pass filter that has desired localisation in space and frequency, we also know the complementary low-pass filter. This is given by the mother wavelet and calculating filter banks is a question of finding this wavelet function with desired properties. For image coding purposes, filters that are short work well and larger kernels are usually not worth the extra cost in computational complexity.

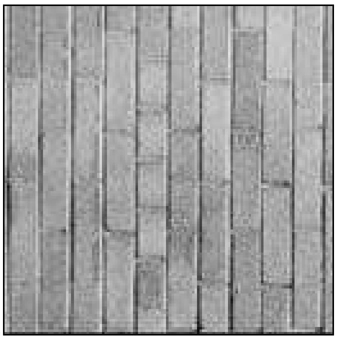
Use of wavelet decomposition in separation

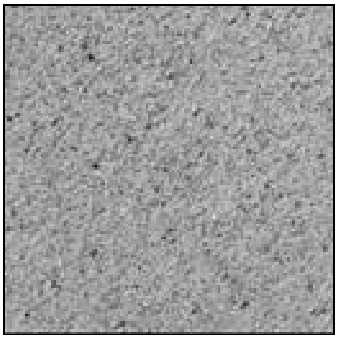
As shown, there are several ways to decompose an image. The same location on every tile in the decomposition is related to the same spatial neighbourhood in the original image. The tiles also form a basis for the image. The different basis functions represent separate frequency bands and to some extent, local orientation [6]. An extraction of carefully selected coefficients will perform separation in both the spatial and frequency domain at the same time.

5. Experiment separating a highly textured image

Commonly used images that are highly textured

These highly textured images are made by the famous photographer P. Brodatz. His photographs of natural textures are as widely used in the texture community, as the Lenna image is in other related research areas. The book is now in print again. VisTex is a public database with similar images.



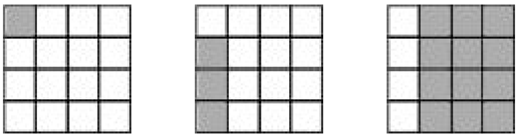


Typical Brodatz' photographs of a brick wall and beach sand. These highly textured images are often used and there are many more in his series. The original pictures are digitised to 512x512 pixels and 8-bit greyscale.

The image to the left shows a brick wall and the other is beach sand. The wall is, on a larger scale, a texture in itself as well as the individual bricks have textured faces. It is all a matter of scale. In this experiment, we will separate the structure of the grid pattern that the blocks make, using wavelet packet decomposition. As mentioned before, the wavelet functions associated with the individual tiles have band-pass properties and are orientation sensitive. By identifying these tiles and separating those into a new basis, separation is achieved.

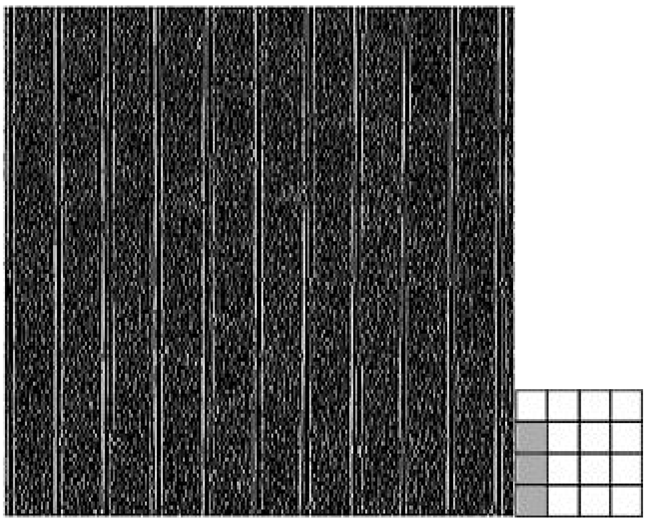
Separation of the tiles

In this experiment, all of the coefficients in a tile are moved to the new separated basis. Moving single coefficients of other tiles is a promising strategy, but on homogeneous images like Brodatz, this is not necessary. Transferring parts of the coefficient in the tile are on the other hand a better idea.



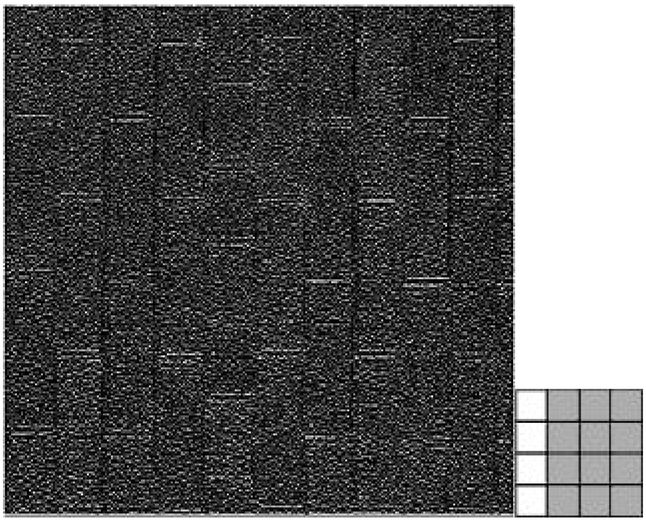
The two-level wavelet packet decomposition is separated into three new bases. The white is void. Analysing the wavelet functions, or rather the corresponding FIR-filter, a separation based on the orientation is expected. The first shows the low-pass tile. This must be treated by itself since almost all of the image energy content is in this band. It is also the only band with a mean value not equal to zero.

The reconstructed image from the three different decompositions contains features that can be described as texture.



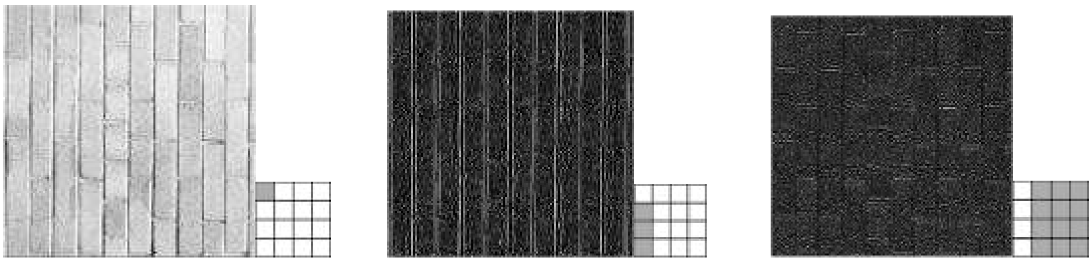
The reconstructed image using a separated basis with only the packets shown in the figure. Since wavelets have a mean of zero, the pixel values has been greyscale coded using their magnitude. The contrast has also been boosted.

Using only coefficients from those tiles that have wavelet functions that capture vertical variations it is possible to separate the structure. The reconstructed image has a mean around zero and this poses a problem when trying to display it as an image. It is necessary to adjust the range and even the contrast. The three selected tiles represent two different scales, but the same direction.



Using some of the other tiles, this image captures the horizontal image content and the details.

The other twelve tiles represent both the horizontal image content and the details of the bricks.



If these three images are added together, the original image is recovered.

The image is now divided into three components that have separate orientation and frequency content. They form an additive decomposition. It becomes clear that they together form a basis when inspecting the respective coverage of the wavelet tiles. The first can be considered local mean in non-overlapping 4x4 neighbourhoods.

At this stage, no real separation has been made since the whole tile, and not only parts of it, where transferred. When doing so, the stack is a convenient representation. The principle of transferring coefficients remains the same.

A way of selecting which coefficients to transfer is by using selection masks created from features in the image. Binary selection masks transfer the whole coefficient.

6. Experiment using selection masks

Wavelet based image coder using binary selection masks

The highly distorted reconstruction of a wavelet-coded image of the Lenna image shown on page 23 is not an attempt to construct a new type of image coder, but merely to illustrate some basic principles. These involve the multi-resolution property of wavelets and *selecting coefficients using binary masks*. Still, a reduction of the data contents is 26 times (0.31 bit/pixel) without any entropy coding.

The use of binary selection masks is the most fundamental concept. The masks themselves are produced using a feature extraction scheme to capture edges.

First, the multi-resolution analysis is clearly shown in the variable block size. This is prominent in the upper left corner. The wavelet of choice is the ultra-simple Haar wavelet. This wavelet has extremely appealing properties, but one drawback is the introduction of block effects when used on the larger scales. For representation of details (eyes) and high contrast textures (hair) it actually might enhance the visual impact quite like an so called *unsharpen mask filter*. The non-overlapping property eliminates the border padding that all other wavelets need. Experiments using a smoother wavelet to reconstruct the coefficients from the two largest scales produced a visually more appealing image with much less prominent false edges. This supports the idea of using the stack representation and a mix of basis functions.

Secondly, the use of masks based on a specific feature in the image, control the extraction of coefficients into a blank basis stack. In a sense, the texture "portrait" has been separated using the feature "edge", supported by a general idea on how the human vision system works. What is left is mostly image content regarded as irrelevant and the reconstructed image holds the more relevant content.

Original Lenna image and the binary selected mask on its largest scale. Only the coefficients whose reconstruction filter-kernels overlap the mask are kept and the remaining 89% are ignored.

Binary selection masks

The use of selection masks is a fundamental step towards data-irrelevancy-reduction and can be used to separate an image. This separation is made in the wavelet domain. The JPEG Q-table can be viewed as a fixed form of mask, but is more tied to a signal model. The mask used here is not fixed and is one of the more basic feature-extraction schemes. The binary mask is produced with an *edge detecting Sobel operator* followed by a threshold of the filter responses magnitude. From this mask, all the smaller masks are made. On the smaller scales, the line widens and thereby represents a larger and larger area of the original image.

Since a line drawing of a person often is sufficient for recognition, it also ought to point towards the relevant, for human vision, regions in the image. Using the edge map to mask out only the wavelet coefficients in these regions, a successful reduction of the irrelevant image content is achieved.

The selection masks at several scales. Since they are binary, an AND-operation is all it takes to mask out the desired coefficients. The masks are independent of each other. Sharp edges are present at all scales, but other features are only present in some. This can be exploited in texture separation.

On each scale, each wavelet coefficient corresponds to a region of a particular size. The table shows how many coefficients that are forced to zero:

```
2x2   92% zeros
4x4   84%
8x8   74%
16x16 59%
32x32 39%
64x64 17%
```

In a total, 233324 of 262144 elements are forced to zero. This equals 89%.

Quantisation to a bitstream

As a third and final step, the coefficients are quantized to three bits using an ad hoc four-times-standard-deviation *uniform quantizer* without a dead zone. The visual impact of this is very small and indicates that the accuracy in the magnitude of the filter response is not a critical issue. Shapiro [A15] concluded that the sign certainty is critical.

Using this crude scheme, a *data reduction of 26 times* is achieved. Note that no entropy coding is done since it is not obvious how to scan the coefficients. Shapiro's zerotree idea [A15] is hard to generalise to wavelet packets. The matrix of coefficients is simply scanned row-wise.

The image of Lenna coded and reconstructed using wavelet coefficients selected by a binary edge map. The Haar filter is used, which introduces false edges on the larger scales.

The reconstructed image is highly distorted, but considering that it only use one tenth of the coefficients and the remaining coefficients are set to zero, the result is not bad. The obvious fine-tuning that can be made is to use smoother wavelets on the larger scales and the highly localised Haar wavelet on the fine scale levels.

Independent selection masks

The masks at different scales are independent of each other. In the experiment above, the masks are created using an edge-detecting filter. Edges are present in all scales making the different masks quite similar. If another feature was of interest, like vertical lines, a special mask for the tiles with that direction and resonance-frequency could be used. It should also be noticed that to create these masks the techniques used are not based on the wavelet decomposition itself. For this, the wavelet coefficients themselves are not a good enough tool and methods using for example quadrature filter and tensor representation are superior [6]. The *Sobel-operator* is a classical and efficient method of finding edges, but operates on one scale only.

To find regions with texture is a very complex task. It is a hard problem to find something that lacks a definition like texture does. A discussion is found in chapter three. If one or several properties can be isolated as characteristic for the texture and then measured using whatever method, the task of automatically making the masks is within reach.

The area of the image is limited and some parts of it are probably not highly textured. As a working model, an inverse definition of texture as what is not texture can be used. Successful segmentation and separation algorithms can be developed [A6]. If a region is homogeneous on several scales, it is probably a texture. A sharp edge or a line would violate this test condition.

Once a region is identified as a textured region, an analysis of the local orientation and the local frequency points out the relevant tiles in the wavelet packet representation. The mask can then be created.

Combining the methods to reach further

So far, experiments using the general ideas of selection masks and the selection of only certain wavelet packets have been demonstrated by themselves. The selection of only certain wavelet packets using the Brodatz brick wall image in chapter five and the selection masks on the Lenna image in this chapter. The goal is to integrate them and pushing further with help of the generality of the stack representation. This is done in chapter nine, but while still on the topic of selection masks, a different form of mask are discussed. These are the non-binary selection masks.

7. Separation of a textured region

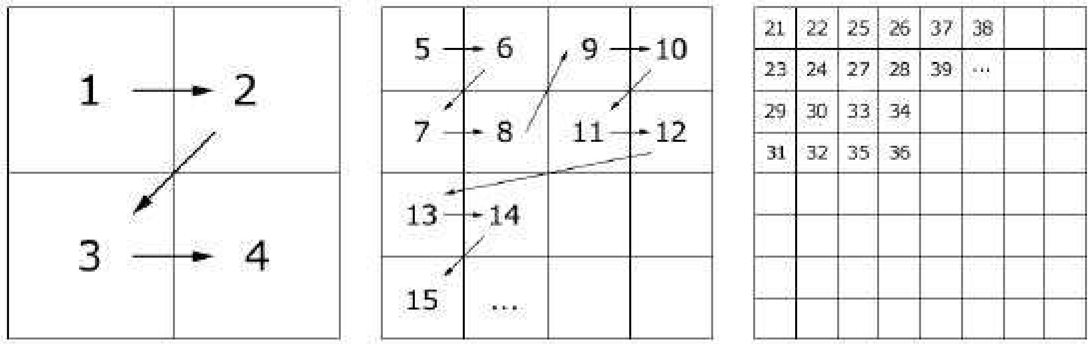
Transferring coefficients selected by masks

By using a method of transferring both the coefficients and using selection masks to guide the selection will give real separation. This scheme is operating only within a subset of the tiles in the representation. The Barbara image shows a girl in striped clothes and a checked tablecloth. These represent textured regions and we shall separate the striped texture on her leftmost leg.

The Barbara (512x512) image to the left, and a close-up of her legs. The striped clothes are here considered textured regions that shall be separated.

Standard way of labeling the tiles

There exist several ways of labeling the tiles. The tree structure presents many possibilities and there is no obvious way that is better than the others are. One method is based on a number-pair where the first is the level of decomposition and the other pointing out the tile in that level. This method still suffers from the drawback that there are several ways to point out the tile. In dyadic wavelet decomposition, there are only four tiles at each level so it still feasible, but expanding it to wavelet packets is cumbersome.



The conventional way of labeling the tiles. The underlying tree-structure is conserved in the labeling and given the number it is straightforward to calculate the level and position of the tile.

In wavelet packet decomposition, the conventional way of labeling the tiles is by running number. These start at the first level. On the next level each parent is given its children labels in order. The figure shows how this is done. Typically, the tile represented by number one is the approximation tile, two is horizontal, three is vertical and finally the detail is in number four. This relation is the same in 9, 10, 11 and 12. The order has nothing to do with the frequency as it sometimes is assumed in the literature [A3].

Another way to represent the decomposition is by a tree. This representation is more suited for one-dimensional data since it only divides in two instead of four.

Wavelet packet decomposition of Barbara

Before proceeding with the separation, the image is decomposed as shown below where we see both the decomposition space and the corresponding tree. The wavelet is Haar. An initial experiment used db4, but the result was only marginally better visually than using the much simpler Haar wavelet. The coefficients are visualised as grey-scale coded by their magnitude.

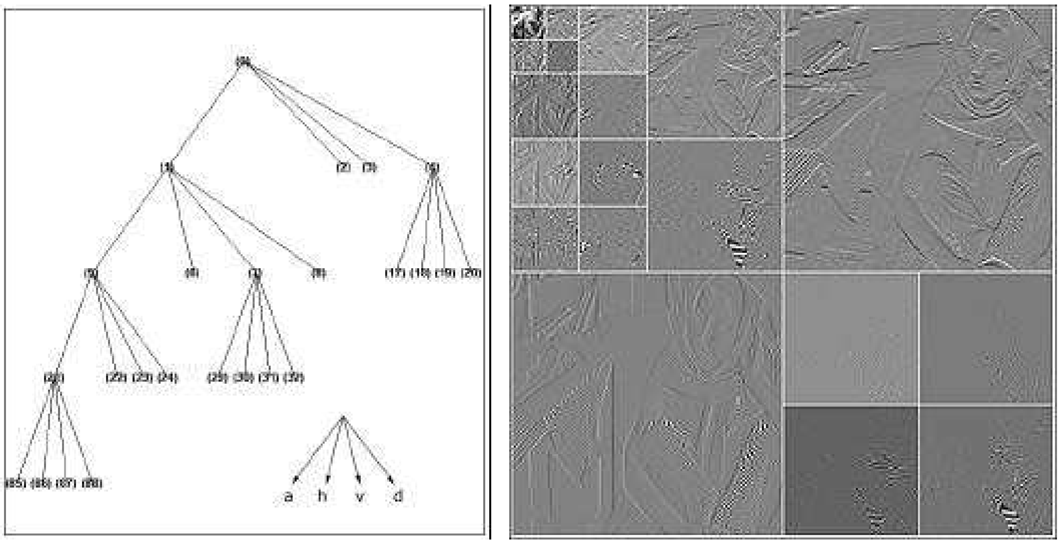
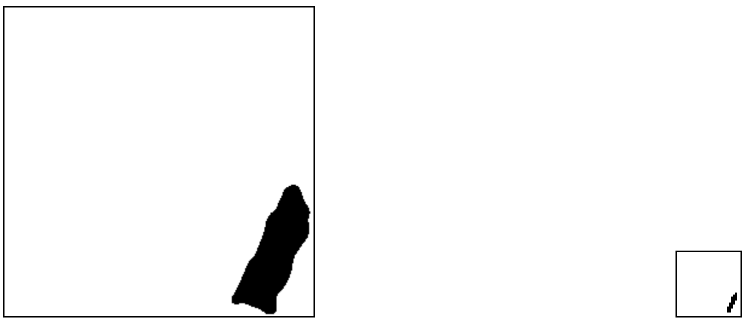


Figure shows a wavelet packet decomposition of Barbara. Both representations show the same tree. Wavelet coefficients are grey-scale coded by their magnitude. (a) approximation, (h) horizontal, (v) vertical and (d) detail.

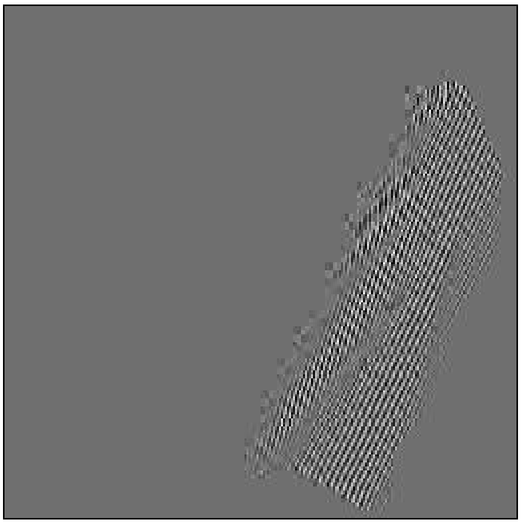
To perform the separation, selected coefficients from tile 3, 8, 17, 19, 24, 30 and 88 is transferred to a new wavelet packet basis. These tiles or packets hold the image content with the orientation of the stripes at several scales. The tiles or packets where chosen based on an analysis of the filter kernels that are related to the tiles. The primary requirement was orientation and secondly the scale. Fourier analysis and visual inspection of the filter kernels assisted the manual pick of tiles.

No selection is made from the low pass tiles since only the texture and not the overall structure was to be separated. The coefficients in tile 3 span a neighbourhood of 2x2 pixels and tile 88 the much larger region of 16x16 pixels. This will give separation with smooth fading along the borders of the region. A set of selection masks is needed. These are produced manually.



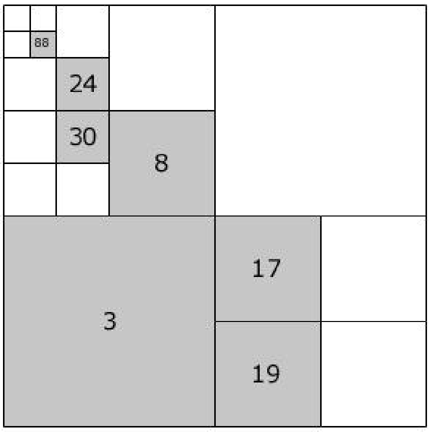
Two of the masks used in the separation. Original size is 256x256 elements and the smallest is 32x32 elements. Five masks are needed.

The masks all refer to approximately the same region but on the coarser levels, the region is narrowing into a line. The selected coefficients are transferred to a blank basis and reconstructed and so are the remaining coefficients. The two images can be added to form the original image. The loss when converting back to eight-bit representation from floating-point is handled. This is done by adding the rounding-off error from the texture image (left image below) to the remaining image (right image below). The sum of the two images then form the original image. The rounding-off error is very small and can not be detected visually.




Result from the separation. The striped leg to the right is separated from the initial image. To the right, the difference in detail and structure can be observed by comparing the texture on the right and left leg.

A separation of Barbara's leg is the result. The region is not cut out sharply as most segmentation methods do, but the detail and the structure of the fabric are separated into a new image. The remaining image still contains the overall structure and the lighting of the scene is conserved. The smooth transition does not introduce false edges, which the human vision system is sensitive to and tends to notice. This controlled blurring might also boost compression ratios in conventional coding schemes as JPEG, but also EZW [A15] or SPITH [A16]. They are all based on a signal model prioritising low frequencies, often called 1/f processes or decaying spectrum.



The tiles used in the separation. Note that tile 17 is the tile with the highest centre frequency and not tile 20 which is a common belief. This is an aliasing phenomenon due to the cascaded use of the same filter.

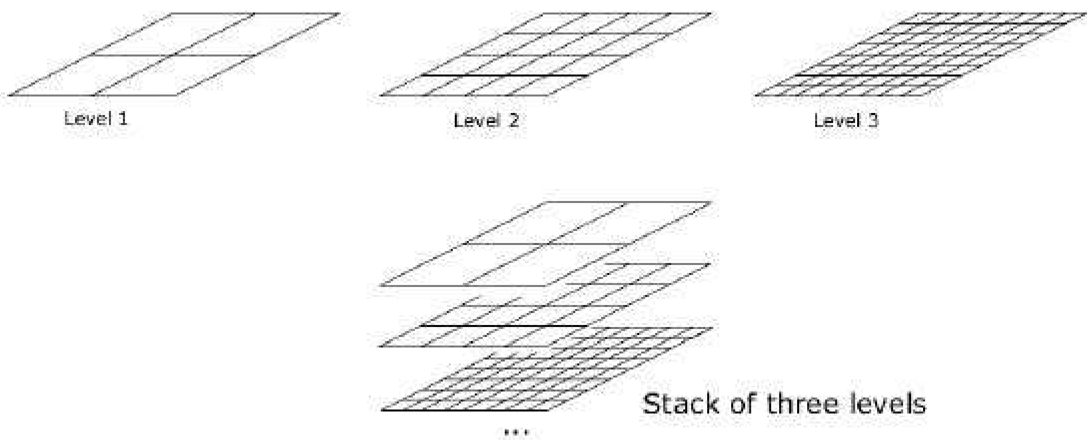
The successful result is a combination between the choice of the tiles, i.e. the filters, and the construction of the masks. Since the masks narrows in on the coarser levels, the borders become smoother than if this mask modification had not been done. The tiles or packets where chosen based on an analysis of the filter kernels that are related to the tiles. The primary requirement was orientation and secondly the scale. To separate the texture of the other leg another set of tiles has to be picked since that texture has another orientation.

8. Stacks, a suitable representation for separation

Introduction of a new representation form

By using a collection of wavelet packet decompositions and selecting a mixture of them, a new basis is produced. The collection is the so-called stack. The initial over-representation is related to frames [6] [A7], but the connection between different levels of decomposition has strong links to the parent-children relation found in the concept of zerotrees [A15].

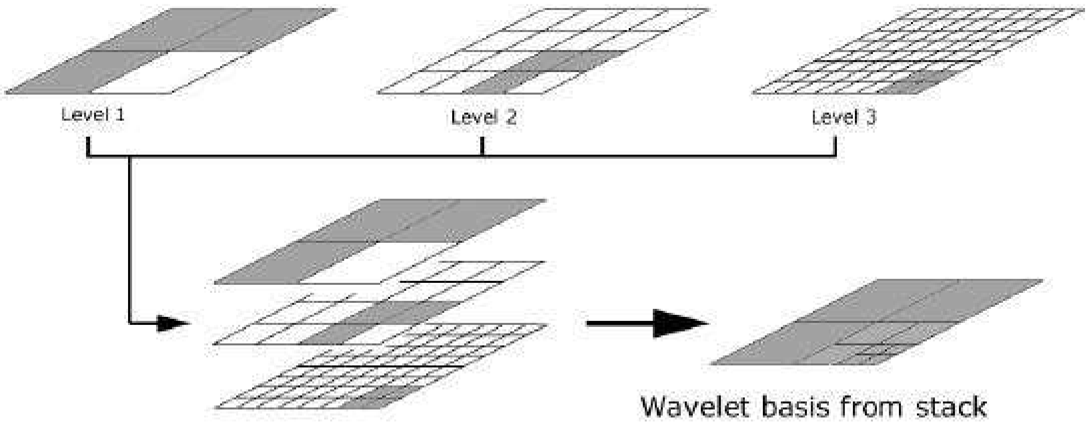
The stack consists of several layers and each layer is a complete decomposition. A Haar wavelet packet stack contains layers of increasing levels of decomposition. Using functions that are non-overlapping, as the Haar wavelet, is preferable. Even two-dimensional DCT of several sizes can be used and stacked.



A basic stack consisting of three layers. These three layers are discrete wavelet packet decompositions. From this stack, a new basis can easily be extracted. The original image can also be added to the top.

The figure shows a basic stack. The dyadic scaling is shown and is in some applications expanding too fast. Astrid Lundmark at ICG, LiTH has proposed other scaling ladders, an advanced topic that requires development of new wavelet families. A problem with the dyadic scaling is that the wavelet functions quickly start to span large areas of the original image and eventually span larger areas than the entire image. Once there, further decomposition is meaningless. Before that, the functions spill over the image's edges and complicated border padding has to be performed. To avoid this highly localised functions can be used.

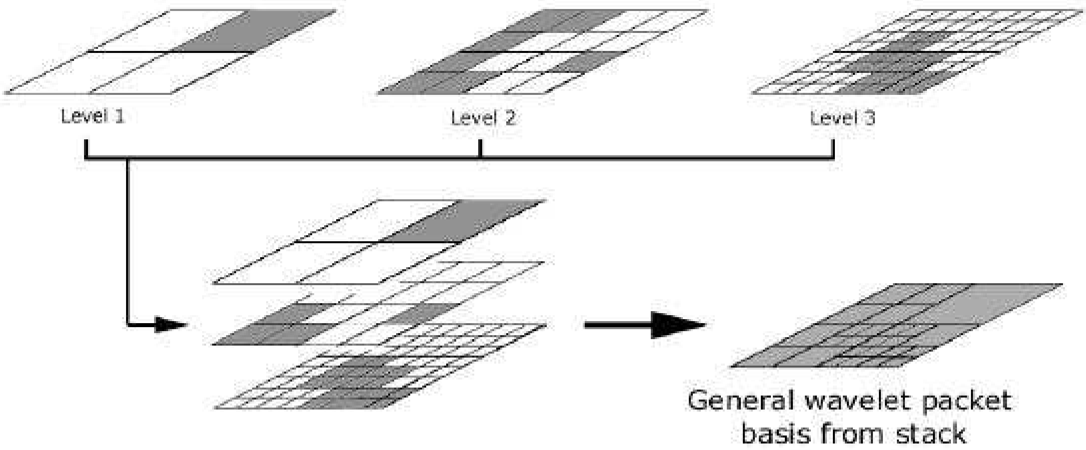
A new basis can be extracted from the stack. If we use the Haar wavelet, that is non-overlapping, this process of extracting a basis becomes straightforward. If the selection of tiles cover the whole area, the basis is complete. Many combinations can be made, but there are an even greater number of impossible selections. The issue is related to quad-tree segmentation and representation [A12].



After forming a stack, a basis can be extracted. In this example the wavelet basis is formed. Looking on the stack from above, all the area is covered and only once. This gives a basis. Other selections can be made.

On every level in the stack, each coefficient is connected to a location in the original image. This relation is trivial and easy to calculate. The corresponding wavelet function also spans different areas depending on the level. Along with the coefficients value, the orientation and resonance-frequency form something that can be viewed as a texton [A11] [6]. From a signal-processing viewpoint, the triple is related to some local properties. Local orientation, local frequency and contrast have connections to the particular wavelet packet and the coefficient value.

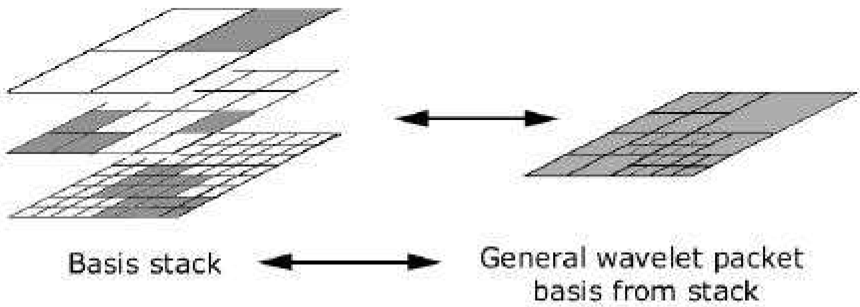
Much of the work found in the literature tries to use the wavelet as a feature-extracting tool. The results are mixed. Most successful is the wavelet used for discriminating textures from each other [A1]. To examine local orientation and local frequency (phase) the wavelet is a too rough tool and methods using quadrature filter and tensor representation are superior. Advances in this field of science are mainly done in computer vision and machine intelligence. This suggests that the feature-extraction should be done with other methods than those that use the wavelet itself. The result, however, from this extraction can be used to steer the separation or compression.



From a three level Haar stack, many basis representations can be made. This example illustrates how to construct a general wavelet packet-basis from the coefficients in the stack. White is void.

Basis stacks from basic stacks

It is possible to build a basis stack from the basic stack. Note the difference between basic and basis. In fact, the extension is simple and in the example above it is just a matter of not deflating the stack to one level. This is shown below.

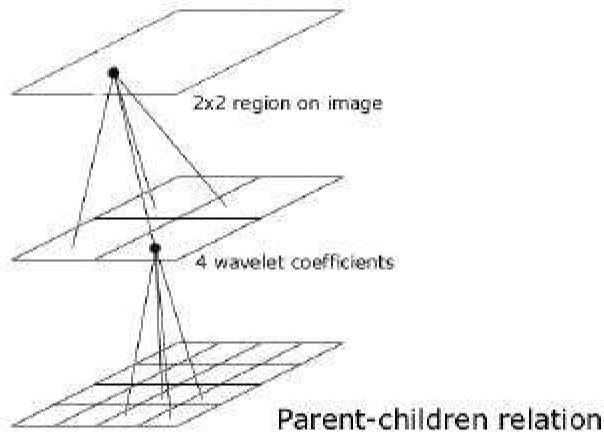


A basis stack. The corresponding, ordinary wavelet packet basis is shown to the right. All the white tiles are void or set to zero so the data contents are identical.

The basis stack contains exactly the same data as the ordinary basis, which could be expected. The representation offers some advantages in generality when used for separation. For this purpose it is convenient to use several basis stacks that are additive. This could for example be done by creating an identical copy of the stack and dividing the coefficients by two. When the two are added together and deflated to one level the more common representation is retrieved. From this, it becomes obvious that the basis stack can also be constructed from the general, wavelet packet decomposition. The use of two stacks is the final product of the separation operation. As the process evolves, coefficients are moved from one stack to the other.

Transferring coefficients between different levels in the stack

Within a basis stack, coefficients can be transferred between the different levels. By doing this transfer, the idea of the tiles covering the whole plane to form a basis collapses. The representation is still complete and no information is lost.



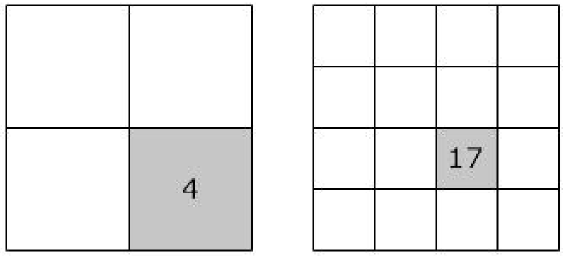
There is a parent-children relation between the levels in the stack. Each coefficient is also related to a spatial location in the original image. In level two each image pixel has a relation to sixteen wavelet coefficients.

Each wavelet coefficient is related to a region in the original image. In the case of the Haar wavelet a 2x2 region is uniquely represented by only four wavelet coefficients in the first level of decomposition. On every level, it takes four coefficients to represent one coefficient on the upper level. Further down, the related region per coefficient becomes larger.

By exploiting this parent-children relation, it is possible to transfer single coefficients down to four new ones or gather four and move up one level. This can only be done in one way if the whole content in the source level is transferred. What is less obvious is the possibility of transferring parts of the coefficients. A way to analyse its consequences is to alter one coefficient and see how the others must be compensated. Whatever is altered inside the stack, the original image must remain the same. This is a complicated operation with many degrees of freedom.

Separation using stacks

Until now, the flat wavelet representation has been sufficient for all the examples and experiments. When picking the tiles of interest within a stack a collision might occur. The feature to be separated can have properties that call for an impossible combination of tiles in the flat representation. The analysis might point towards both tile 4 and tile 17, which overlap.



Tile 4 and tile 17 picked to be used in the selection. This collection of tiles is impossible to represent if not using stacks. When forming the basis stack only a fraction of the coefficients in tile 4 or tile 17 can be transferred. The tiles have the same centre-frequency but different support.

When forming a basis stack from the two levels in the basic stack above, only a fraction of the coefficients can be transferred. Since the two tiles have the same centre-frequency and orientation, this transfer can be used for energy compaction into as few coefficients as possible. If a neighbourhood of 2x2 coefficients in tile 4 is of approximately the same value, they can be lowered to tile 17. The magnitude of the corresponding coefficients in tile 18, 19 and 20 will then become very small. Four coefficients will be represented by only one. The representation is redundant.

To remain the basis for the image, if only tile 4 and tile 17 are to be used, tile 4 can only contain the information of tile 18, 19 and 20. Either the information from tile 17 is subtracted from tile 4 or it is blanked and built up again by inverse transformation of tile 18, 19 and 20.

9. Advanced selection masks

Non-binary selection masks

Using a binary mask, the whole coefficient is transferred to another stack or another level in the current stack. Using a non-binary mask, only a part of the coefficient would be selected for transfer. Normally this mask would weight out a fraction of the coefficient, which is rather uncomplicated to implement. The analysis of the results becomes rather complicated on the other hand. If the separation is done in purpose of aiding a compression algorithm, normally an energy compaction is desired. The use of several stacks and transfers between them is not in line with that strategy.

Many commercial JPEG encoders have the option of an initial smoothing before coding to boost the compression ratio and reducing the specific compression artefacts associated with the JPEG algorithm. In a sense, this smoothing is a frequency-related decomposition where only the low pass component is regarded.

Selecting the tile and the coefficient in question has been discussed and the criteria for these selections can be described explicitly. When it comes to the question of transferring only a fraction of a coefficient, the reason is not obvious and the decision criteria harder to formulate. For example, this may happen when a very smooth border is desired or to remove lighting effects as shadows by selecting small fractions of the low pass tile coefficients. This is hard to automate in an algorithm.

In chapter 7, non-binary masks are not used directly in the separation process but in an initial rearrangement of the stack with internal transfers between the scales. The separation is then done using binary selection masks in a stack representation.

Non-linear masks

A third form of mask is the non-linear mask. These have not been investigated. The masks are only discussed here in general terms as the next level of complexity.

Median or threshold masks are two suggestions using non-linear filters. The threshold can be adaptive and the median neighbourhood restricted. The result, however, is very difficult to predict from such operations. A fundamental problem with linear filters is that they always give a response [A11] regardless of whether the filter matches the signal or not.

In many fields of science, the use of approximation is commonly used to estimate model parameters from empirical data. Using only linear approximation, most prediction models fail but as more terms, quadratic cubic etc, is added the curve fit rapidly becomes better. Considering general moments the same principle is found. Non-linear masks might be the method of the future, but the analysis is extremely complicated.

Linear approaches are in this sense considered first-order approaches in analogy with the Spline and moment concept.

10. Literature study

Overview

As a part of the master project described in this thesis, a literature study and gathering of articles was to be performed.

To this date, there is no book on the topic available and most sources of information are found in articles and doctoral theses. For this reason, this part of the thesis serves as a commented list of references.

One goal of the study was to collect the latest and best introduction to wavelet theory, mathematical as well as to practical applications. This was not easy due to the young nature of the subject and most literature is so to say, the first wave and published in the period from the early to mid nineties. It is expected that newer publications manage to refine the presentation and contents.

There seems to be two kinds of wavelet introduction books and the background of the author mostly determines the type. Mathematicians write one kind and signal processing engineers another. Chui's book [1] however, manages to combine the two descriptions into a compact, but still complete presentation. The book is by no means easy and requires a few passes.

The clearest, as well as the shortest, presentation of wavelet theory is found in the doctoral thesis [A1] and [A10], of which the latter is the slightly better one. Gert Van de Wouwer has written a lot on textures and segmentation. His website on texture analysis is no longer maintained. John R Smith's thesis [A9] also contains a presentation of the wavelet theory, but lacks the mathematical presentation that the theory deserves. His work [A9] and [A14] is non-the-less brilliant work and is richly illustrated and exemplified. These references, [A10], [A1] and [A9], form the frontline in texture segmentation to date and a summary is found in [A8] written by Gert Van de Wouwer et al.

The other goal with the literature study was to find out what was the status of the research in texture segmentation using wavelet techniques. Faced with this task, one quickly finds out that the term texture has many different definitions. Equally fast, it becomes obvious that segmentation also lacks a common denominator and the process of narrowing in became time-consuming. Lots and lots of impressive and interesting work have been done in neighbouring research areas and keeping focus is not easy for the curious-minded.

Early in the work with this thesis, the use of the term segmentation became uncomfortable and eventually the wider term separation was adopted. Depending on the underlying proposal for this sort of manipulation, this distinction might be of more or less importance. To avoid confusion and misuse of the common praxis separation is the chosen term. The difference between segmentation and separation is discussed in chapter 3.

The remaining references that have not been mentioned above consist of articles and single chapters in books. Some point to dead ends and others to related areas, but they all contain ideas that are worth a closer look. If the comment seem harsh, it is by no means to discredit the authors.

In the appendix, a summary of a wavelet packet based texture image coder used to store fingerprints is presented. It is called WSQ and since it is designed exclusively for use on highly textured images as fingerprints, we found it appropriate to include it in this thesis.

List of references with comments

The comments are with respect to Image Coding Group's work with wavelet packets on textures, but might also help the reader in selecting literature for their purpose. Each of the references has its own reference list that ought to be of interest. The numbering of the books and articles are the same throughout this entire thesis. Much of the information is available online on the Internet, but due the dynamic nature of it, most universal resource locators are omitted.

Books

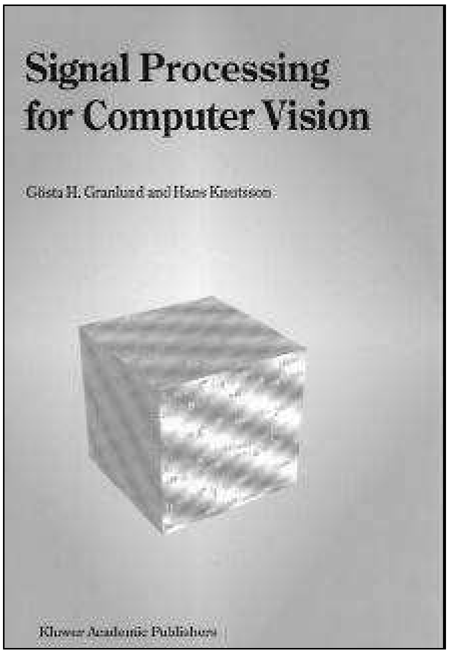
[1] An introduction to wavelets
Chui, Charles K.
ISBN 0-12-174584-8, 1992
This book contains a compact and good mix of mathematics and signal analysis. This is the only book recommend as entry-level literature. Especially chapter 3 on short time fourier transform has a good introduction to time-frequency analysis and it also covers Gabor window (transform). This transform is optimal with respect to the uncertainty principle and has many uses and connections to wavelet based techniques in texture analysis and computer vision. The book is complete and has the depth into the subject that can be expected.

[2] A friendly guide to Wavelets
Kaiser, Gerald
ISBN 0-8176-3711-7, 1994
It is not friendly at all. Chapter 2 and 3 might be of interest as an overview to the wavelet transform. The author has published other related material of more interest.

[3] Adapted Wavelet Analysis from Theory to Software
Wickerhauser, Mladen Victor
ISBN 1-56881-041-5, 1994
A book with a good hands-on approach which is very useful if you intend to implement computer routines yourself. Chapter 7 deals with wavelet packets and the cascaded use of filters. The best-basis algorithm is discussed in chapter 8. Chapter 10 deals with time-frequency analysis. The appendix includes some orthogonal QF coefficients as well as a couple of biorthogonal QF coefficients. Some editions of the book contain errors in these filter coefficients so double-check your filters if your source is the ones found in this book.

[4] Wavelets: A tutorial in Theory and Applications
Chui, Charles K
ISBN 0-12-174590-2, 1992
A hardcore mathematics based book almost free of figures. It covers wavelet transforms and filter banks in great depth. However, at the end of the book, there are some articles and especially one of them presents a rather new idea. It is called the Second generation compact image coding with wavelets. Textures, defined as the error image, are coded separately and added. The overhead is too costly to give a good compression ratio, but the underlying idea can be useful for segmentation schemes. Feature extraction is a part of the algorithm and a similar concept is used in this thesis in chapter 7.

[5] Video Coding, The Second-Generation Approach
Edited by Luis Torres and Murat Kunt.
ISBN 0-7923-9680-4, 1996
It is a very novel set of ideas presented and the 5:th chapter, Region oriented texture coding, by Michael Gilge is of interest. Texture is in this text defined as the contents of a region. A good description of generalised moments of which the Fourier basis happens to be one is enlightening reading. The LMS approximation and transform coding is shown to be equivalent. The use of polynomial approximation is in fact a low pass filtering operation. In a method description using the concept of region of interest, ROI, the overhead created by the shape information of the arbitrary shaped region proves to be quite costly. In Shapiro's article [A15] on EZW it becomes clear that this overhead, if any at all, can be tolerated if an extremely compact representation is to be achieved.



Cover of the book Signal Processing for Computer Vision [6]

[6] Signal Processing for Computer Vision
Edited by Gösta H. Granlund and Hans Knutsson
ISBN 0-7923-9530-1, 1995
Especially chapter 13, Texture Analysis by Morgan Ulvklo, deals with topics of feature extraction, discrimination and segmentation. The methods use QMF to estimate local orientation and frequency and these methods are quite robust. The authors are primary involved in computer vision. The techniques can be used to produce selection masks to steer segmentation.

Articles and doctoral thesis

[A1] Performance measures for Wavelet-based Segmentation Algorithms
Navid Fatemi-Ghomi
Surrey University, September 1997.
Doctoral thesis.
As a texture-discriminating tool, the wavelet transform is used on images. The performance is measured for many different wavelets.

[A2] Joint Space-Frequency Segmentation using Balanced Wavelet Packet Trees for Least-cost Image Representation.
Cormac Herley, Zixiang Xiong, Kannan Ramchandran and Michael T. Orchard
Hewlett-Packard Labs, University of Illinois
IEEE Trans. on Image Processing, vol. 6, pp. 1213-1230, September 1997
This article shows an uncomplicated method of segmentation that gives excellent results in terms of compression. Keywords are wavelet-tree, single-tree, double-tree, dual double and space-frequency tree.

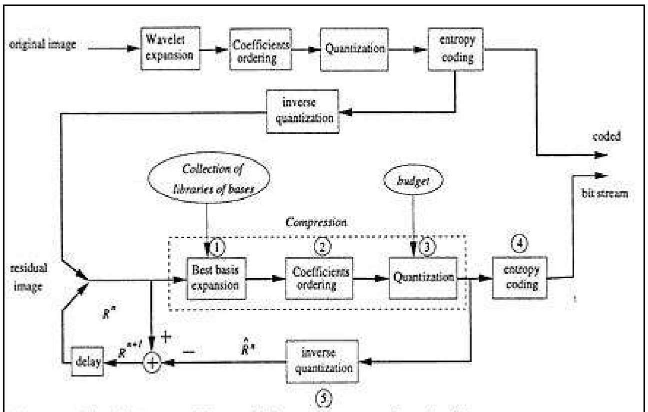


Figure 1. Block diagram of the multi-layered compression algorithm.

Facsimile taken from [A3] showing a flowchart of the multi-layered image compression algorithm.

[A3] Multi-layered Image Compression
F.G Meyer, A.Z Averbuch, J.O Strömberg and R.R Coifman
IEEE Transactions on Image Processing, September 1998
Impressive work and very nice results even below 0.25 bit/pixel. They use a library of different bases, starting with an ordinary wavelet and continue to code the error image recursively using other basis functions. The experiments are made on the Barbara image.

[A4] Wavelet Probing for Compression Based Segmentation
Baiqiao Deng, Björn Jawerth, Gunnar Peters and Wim Sweldens
Mathematical imaging (1993) 2034/266-.
Attempts to use wavelets to find cutting points in one-dimensional data for separate compression of the segments. The result on non-artificial data is horribly bad. The method is clearly a dead end.

[A5] A Robust Automatic Clustering Scheme for Image Segmentation Using Wavelets
Robert Porter and Nishan Canagarajah
IEEE Transactions on Image Processing Volume 5 no 4 April 1996.
Interesting algorithm for automatic true cluster number detection that requires no threshold. The optimal feature selection is ordinary multi-resolution analysis, using the sub-bands energy content as discriminating feature. Local orientation is not considered as a texture feature.

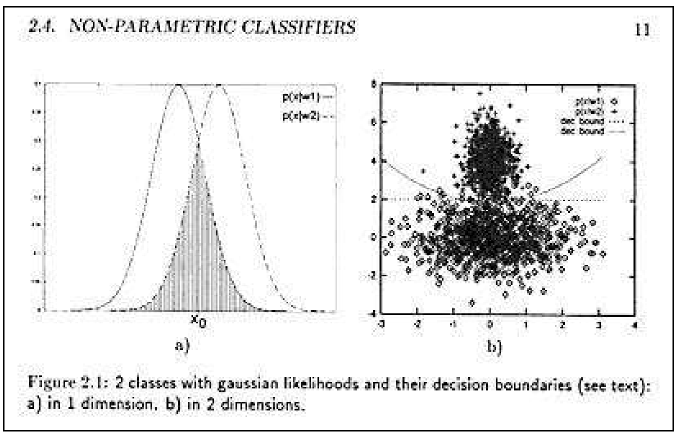
[A6] Texture Segmentation using Wavelet Packets
Yu-Chuan Lin, Tianhorng Chang and C-C Jay Kuo
Mathematical imaging (1993) 2034/277-.
The algorithm uses a fuzzy logic algorithm. Some parameters, alpha and beta, are not explained which makes the examples difficult to assess. The main idea is to use a sort of energy measurement of the sub-bands as features. The method uses thresholds, but does not group the subbands and

thereby losing information of the textures orientation. It contains many examples and illustrations.

[A7] Unsupervised Texture Segmentation using Discrete Wavelet Frames
S. Liapis, N. Alvertos and G. Tziritas
Institute of Computer, Department of Computer Science, University of Crete, Greece. http://www.csd.uch.gr/~liapis/publications.html
European Signal Processing Conference, 1998
This article is very interesting. They use discrete wavelet frames, DWF, to extract a feature vector. DWF are related to WP but instead of sub-sampling the image, the filter is up-sampled. The redundant representation improves robustness and the correspondence between the scales become simple. The features are created from variance of the pixels in an automatically detected homogenous region. Results are good and the simplicity of the algorithm is very appealing.

[A8] Wavelets for Texture Analysis
S. Livens, P. Scheunders, G. Van de Wouwer, D. Van Dyck
University of Antwerp, Belgium
IEE conference Image Processing and Analysis in Dublin 1997
`http://www.ruca.ua.ac.be/~VisionLab/`
Overview of the field of texture analysis using wavelets. Excellent literature list. One of the authors is Gert Van de Wouwer who wrote [A10].

[A9] Integral Spatial and Feature Image Systems: Retrieval, Analysis and Compression
John. R Smith, Colombia University Press 1997
Doctoral thesis
Dealing mainly with automatic texture extraction from large data bases, Smith also discusses the feature extraction from the wavelet subbands. This thesis describes all possible spatial and frequency splitting that can be done using either quad-tree or wavelet decomposition.



Facsimile taken from the doctoral thesis [A10] displaying some test results of a classifier.

[A10] Wavelets for Multiscale Texture Analysis
Gert Van de Wouwer, University of Antwerpen, Belgium, 1998
Doctoral thesis
This doctoral thesis is a collection of independent chapters. The coverage of wavelet theory and texture analysis in chapter 3 and 4 is well done. The chapter on rotation invariant wavelets is interesting for texture analysis application, since a rotated zebra is still wrapped in a skin with zebra texture. Discussion on colour texture analysis is the only one found in this literature study.

[A11] Textons, Contours and Regions: Cue Integration in Image Segmentation
Jitendra Malik, Serge Belongie, Jianbo Shi and Thomas Leung, Computer Science Division, University of California at Berkeley
International Conference on Computer Vision, September, 1999
Discusses challenges in image segmentation and presents a method that combines contour and texture features into the analysis. The method does not use wavelets although the concept of multiscale is recognised. The problem with linear filter responses is formulated.

[A12] Quad-Tree Segmentation for Texture-Based Image Query
John R. Smith and Shih-Fu Chang, Columbia University
ACM 2nd International Conference on Multimedia, 1994
Multi resolution analysis is used on quad-tree segmented images to extract features. The feature space is partitioned to reduce its dimension. This article is a part of [A9]. See also [A2].

[A13] Texture Representation and Synthesis using Correlation of Complex Wavelet Coefficient Magnitudes.
Javier Portilla and Eero P. Simoncelli
CSIC Technical Report #54, April 1999
Impressive work that gives some insight in how to model textures. Their synthesis and extrapolation of textures give very good visual results. The paper is not that relevant for image coding, but helps to understand how texture could be defined and modelled.

[A14] Frequency and Spatially Adaptive Wavelet Packets
John R. Smith and Shih-Fu Chang Colombia University
0-7803-22431-5/95 IEEE 1995
The paper presents a tree decomposition in both frequency and space. The complexity is the same as the double-tree decomposition. The results are 3-5 dB PSNR better than JPEG at low bit rate. The author uses rate-distortion criteria to optimise the tree. See also [A9].

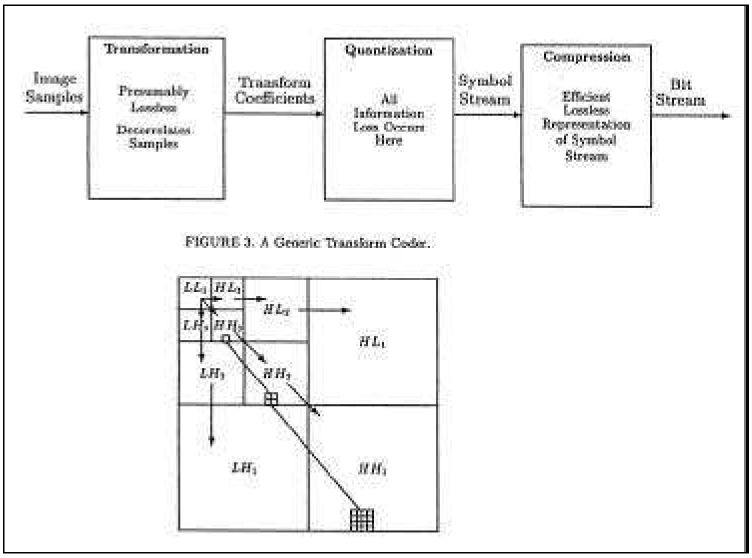


FIGURE 3. A Generic Transform Coder.

Facsimile taken from [A15] showing a generic image coder and the irrelevant zerotree concept.

[A15] EZW, Embedded Image Coding using Zerotrees of Wavelet Coefficients
Jerome M. Shapiro
IEEE Transactions on Signal Processing, pp 3445-62, December 1993
This article presents the EZW image coder. The scanning of the coefficients is done using irrelevant zerotrees. Mandatory reading for everyone involved in image coding.

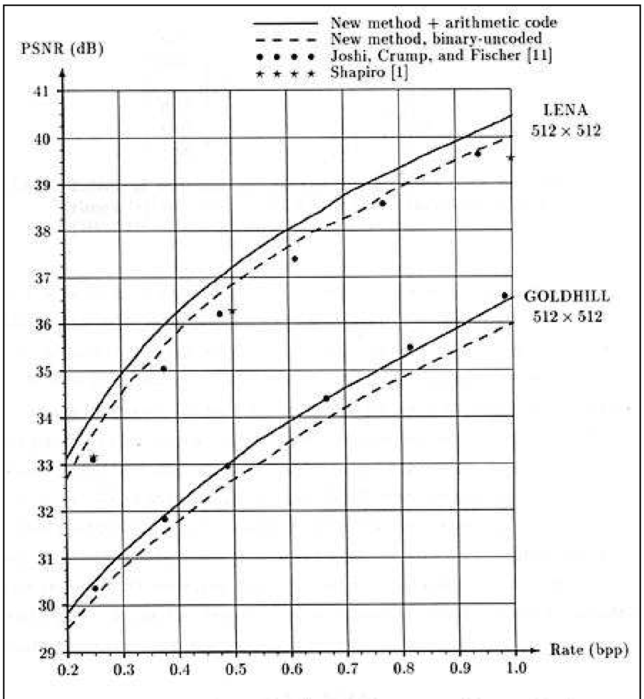


Figure 3: Comparative evaluation of the new coding method.

Facsimile from [A16] comparing EZW and the newer SPITH image coder

[A16] SPITH, A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees
Amir Said and William Pearlman
IEEE Transaction on Circuits and System for Video Technology, vol. 6, June 1996 (originally from 1993).
SPITH is a very good compression scheme. It contains references to [A15].

Classics

[H1] Decompositions of hardy functions into square integrable wavelets of constant shape
A. Grossman and J. Morlet
SIAM J. Math Anal., 15:723-736, 1984.

[H2] Orthogonal bases of compactly supported wavelets
I. Daubechies
Comm. Pure Appl. Math., vol41, 909-996, 1988.

[H3] A theory for multi-resolution signal decomposition: The wavelet representation
S. Mallat
IEEE Trans. Pattern Anal. Machine Intell., vol11(7), 674-693, 1989.

[H4] Ten Lectures on Wavelets
I. Daubechies
Capitol City Press, Montepellier, Vermont, 1992.

WSQ

[WSQ1] WSQ Fingerprint Image Compression Encoder/Decoder Certification Guidelines, January 12, 1999, National Institute of Standards and Technology (NIST).
Found at for example:
`http://www.itl.nist.gov/iaui/894.03/fing/cert_gui.html`

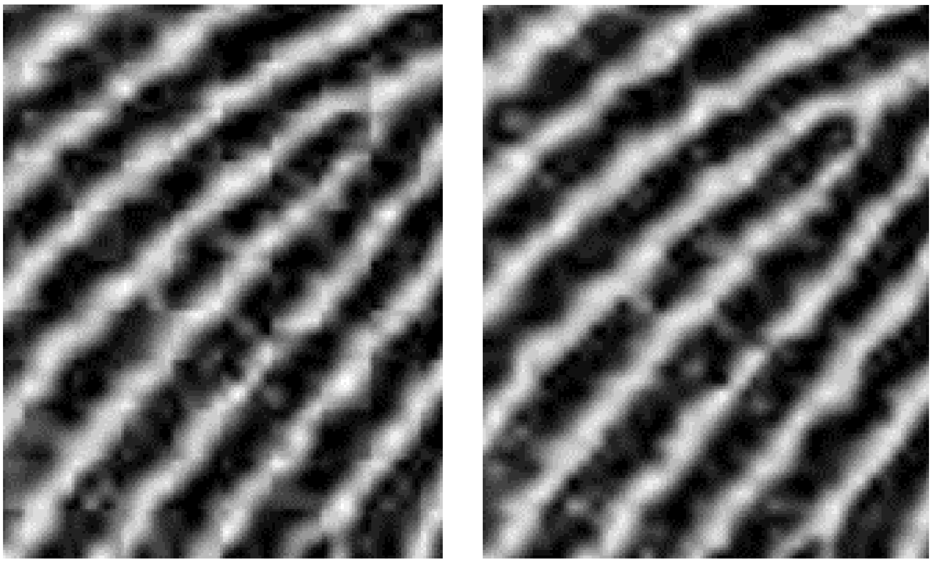
[WSQ2] The FBI/Yale/Los Alamos [W]avelet-packet [S]calar [Q]uantization fingerprint compression algorithm, for Windows 3.1 or higher, by He Ouyang and M. Victor Wickerhauser Washington University in St. Louis.
Executable program file found at for example:
`http://archives.math.utk.edu/software/msdos/miscellaneous/wsqwin/wsqwin.zip`

A. Appendix

The WSQ - file format

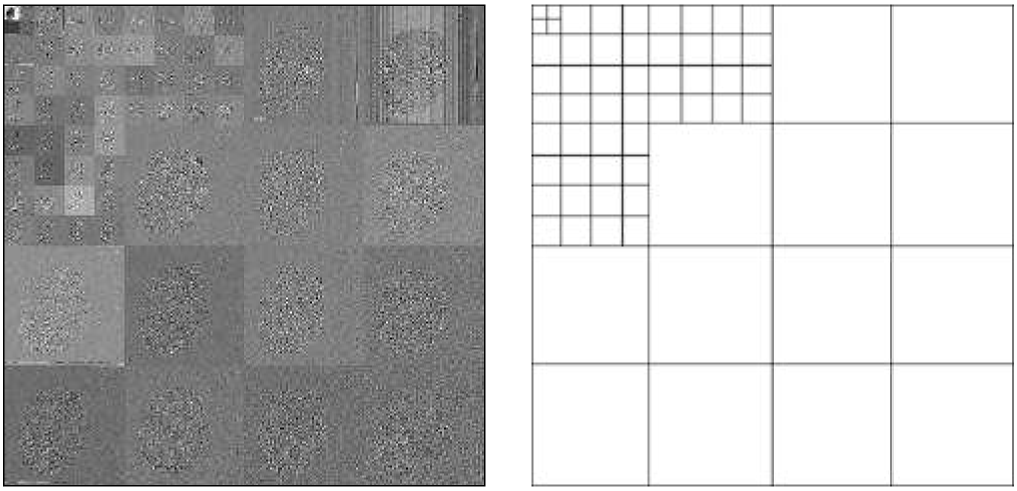
The Wavelet Scalar Quantisation (WSQ) Grey-scale Fingerprint Image Compression Algorithm is a standard for the exchange of fingerprint images within the criminal justice community. The WSQ Specification defines a class of encoders and a single decoder with sufficient generality to decode compressed image data produced by any compliant encoder [WSQ1].

The image format stores greyscale images at a target rate of approximately 0.75 bits per pixel and along with a Huffman coder yields a compression ratio of 15:1. At this target rate, the DCT-based industrial standard compression JPEG starts to introduce block artefacts. This type of artefact can not be tolerated within the criminal justice community for obvious reasons.



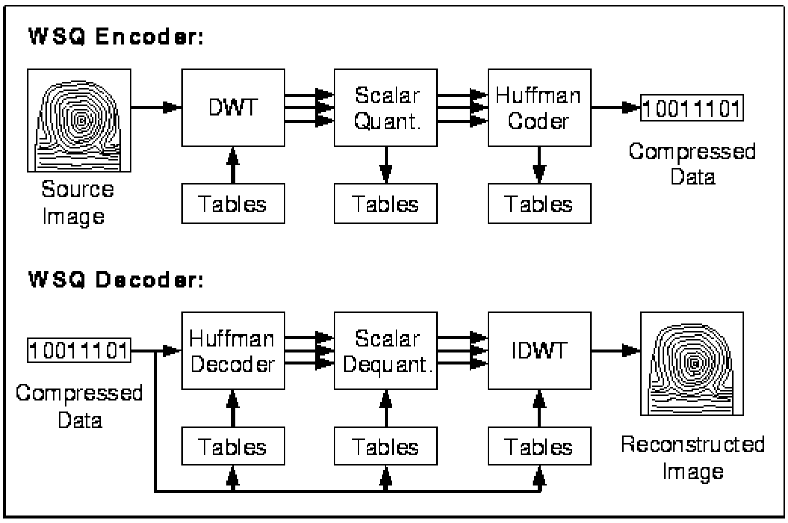
Zoomed 4x JPEG compared to WSQ at a rate of 0.75 bits per pixel. Note the block artefacts on the JPEG coded image to the left.

The use of wavelet decomposition is one way to deal with this artefact since the discrete wavelet transform operates over several scales and not in fixed 8x8 pixel blocks. The decomposition is made into 64 bands that split some of the high-pass band. This is called *wavelet packet* decomposition, and the mid-bands are in fact band-pass filter responses. They are orientation sensitive as can be seen on the example of the 64-band decomposition below. This choice of decomposition tree is probably done with the insight that fingerprints represent a type of texture that has strong resonance frequencies in the mid-bands. For natural objects, the ordinary wavelet tree is a better choice with respect to subband energy.



Example of WSQ 64-band discrete wavelet decomposition and for clarity the grid of the separate subbands.

The mother wavelet used to encode the fingerprint is not fixed by the standard. The filter coefficients must therefore be embedded in the data format together with the tables for the scalar quantisation as well as the Huffman coder table. This makes the WSQ format versatile and open to advances in the knowledge of wavelet filters, or for that matter other subband filters. For an encoder to be accepted for certification it must manage to compress a set of publicly available test images according to specifications in part III of the certification guidelines [WSQ1].



The tables of the wavelet, quantisation and Huffman coder must be embedded in the bitstream.

Today there are many commercial encoders and decoders on the market, but most companies have the compression software as a part of a biometrics system that is sold on a turnkey system basis. One free software package, written in 1993, is available for PC-systems running Windows [WSQ2].

Mladen Victor Wickerhauser is co-author and have greatly contributed to the development of wavelet based applications [3] over the years.

Reference:

[WSQ1] WSQ Fingerprint Image Compression Encoder/Decoder Certification Guidelines, January 12, 1999, National Institute of Standards and Technology (NIST).
Found at for example:
`http://www.itl.nist.gov/iaui/894.03/fing/cert_gui.html`

[WSQ2] The FBI/Yale/Los Alamos [W]avelet-packet [S]calar [Q]uantization fingerprint compression algorithm, for Windows 3.1 or higher, by He Ouyang and M. Victor Wickerhauser Washington University in St. Louis.
Executable program file found at for example:
`http://archives.math.utk.edu/software/msdos/miscellaneous/wsqwin/wsqwin.zip`