Wavelet Networks

Eduardo Martin Moraud (s0896801)

February 6, 2009

Contents

1 Introduction

Many applications of signal processing entail detecting, extracting and classifying specific elements from high-dimensional data. These may be particular sounds from acoustical signals, or shapes from visual scenes, and may most certainly present features that distinguish them from the surrounding (sush as a color or a frequency). A search strategy that concentrates on **transients** would, under these circumstances, enable to easily separate features and underline edges. This is, as a matter of fact, the key idea underlying the wavelet transform [6].

Limitations of classical tools: Classical tools for signal processing such as the Fourier Transform (FT) have interesting properties for emphasising frequential features. However, because of the way they are defined, they are unable to distinguish signals that are stationary from others which vary over time. That is due to the fact that the FT correlates signals with sinusoidal waves ($e^{i\omega t}$) which have *non-compact support* and cover the whole real line. The transform thus yields a global 'mix' of information that makes is difficult to account for local frequential properties.

Figures 1 and 2 illustrate this limitation. The Fourier transforms of two signals with identical frequency components are compared. The first one is stationary: $f = \sin(2\pi f_1 \cdot t) + \sin(2\pi f_2 \cdot t)$ for $t \in [0, 5]$, whereas the second is time-varying: $g_{[0,5]} = \sin(2\pi f_1 \cdot t)$ and $g_{[5,10]} = \sin(2\pi f_2 \cdot t)$. It may be noted that the resulting transforms are identical, and that it is thus impossible at the sight of the spectrum to determine what signal was at its origin.

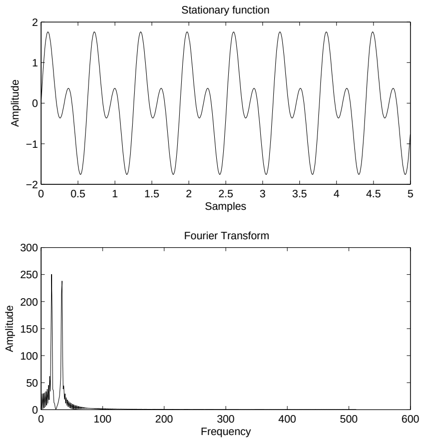

Figure 1: Fourier Transform of a stationary signal $f = \sin(2\pi f1 \cdot t) + \sin(2\pi f2 \cdot t)$ for $t \in [0, 5]$. Two peaks at $f1$ and $f2$ are clearly visible in the frequency domain.

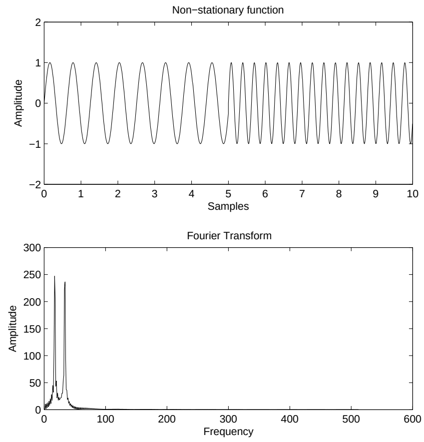

Figure 2: Fourier Transform of a non-stationary signal g defined as: $g_{[0,5]} = \sin(2\pi f1 \cdot t)$ and $g_{[5,10]} = \sin(2\pi f2 \cdot t)$. The transform is identical to that of the stationary signal.

What makes Wavelets different? The wavelet transform has been specifically conceived for evaluating *time-varying frequency information*, and it is especially suitable for highlighting time-frequency properties in that it decomposes functions over test functions $\psi_{s,u}$ that have *compact support* and minimal spread in the time-frequency domain. Figure 3 illustrates the wavelet transform for the aforementioned signals, and emphasizes its utility for detecting frequential transients.

This property of the WT may also be applied to 2D functions – images – in order to detect edges and contrasts. The example in Figure 4 shows the possibility to point out contours, and it confirms its applicability for shape tracking or face recognition.

• In the remainder of this review, I will thus present the foundations of Wavelet Networks and their applications in a variety of domains. Section 2 starts by clearly defining the Wavelet transform and the neural networks that are derived from them. Section 3 then presents an overview of different applications, both for engineering and computer vision.

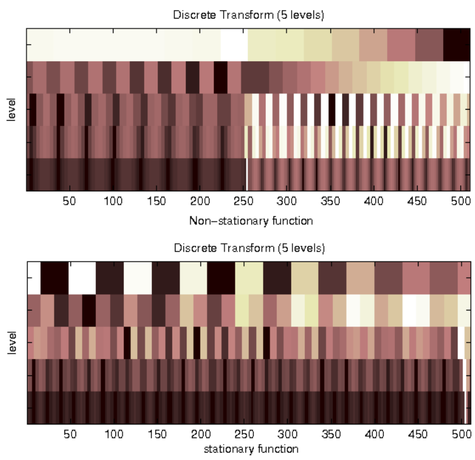

Figure 3: Wavelet Transform of the previous signals: Non-stationary (top) ans stationary (bottom).

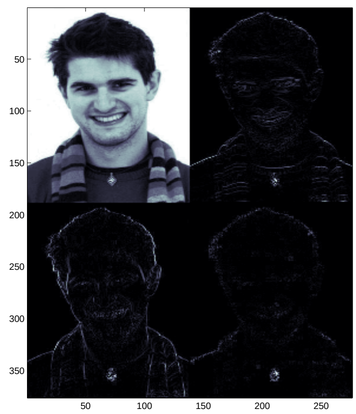

Figure 4: 2D Wavelet Transform of a picture with 1 coefficient.

2 Formal Description

2.1 The wavelet transform, a tool for decomposition

The Wavelet transform is based on the same principles as the Fourier transform, but relies on a finite-support test function – the wavelet – which acts as a localised filter of the original signal, making it possible to iteratively analyse *windowed* parts of the signal and extract space-dependent frequential information.

Each wavelet is derived from a zero-mean 'mother' function ψ (Fig 5) through two linear transformations: (1) Dilatation by a scale parameter 's' and (2) translation by 'u'. These parameters determine the width of the window and hence define the *resolution* of the transform, as illustrated in Figure 6:

$$\psi_{s,u}(x) = \frac{1}{\sqrt{s}} \cdot \psi\left(\frac{x-u}{s}\right)$$

The wavelet transform of a function $f(x)$ – for a given resolution (s,u) – is then given as:

$$\mathrm{WT}_{s,u}\{f(x)\} = \int_{-\infty}^{+\infty} f(x) \cdot \frac{1}{\sqrt{s}} \psi^T \left(\frac{x-u}{s}\right) dx$$

which, in the Hilbert space $\mathbf{L}^2(\mathbb{R})$, corresponds to the inner product $< f, \psi_{u,s} >$, i.e. the projection of the function in the direction of vector $\psi_{u,s}$.

As a matter of fact, the set $\{\psi_{u,s}\}_{u,s\in\mathbb{R}}$ represents a *basis* of $\mathbf{L}^2(\mathbb{R})$ over which any given signal with finite energy may be **decomposed** into its different frequency bands:

$$\tilde{f}(x) = \sum_{\forall u,s} < f, \psi_{u,s} > \psi_{u,s}(x) \tag{1}$$

For a finite number of components N, the function $\tilde{f}(x)$ is an **approximation** of the original function in the space spanned by the N vectors $\{\psi_{s,u}(x)\}$. Interestingly, it may possible to ensure that the approximation is as close as desired to the original function by selecting the appropriate size for the subspace [6], i.e.

$$\forall \epsilon\ ,\ \exists\ N \text{ so that: } \parallel f(x) - \sum_{n=1}^{N} < f, \psi_n > \psi_n(x) \parallel < \epsilon$$

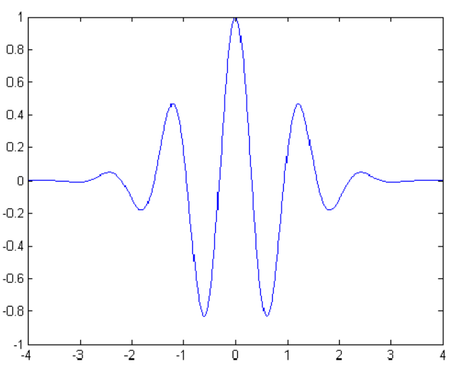

Figure 5: Example of Morlet wavelet mother function.

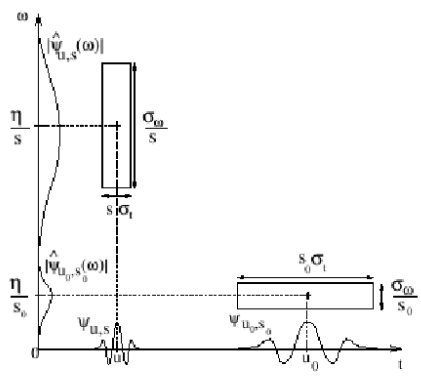

Figure 6: Waveform support in time ('windowing') and corresponding frequential resolution (Extracted from S. Mallat 1999 [6])

This property automatically leads to the idea of **filter banks**. These combine parallel decompositions over different resized wavelets in an attempt to approximate non-linear signals while ensuring that certain desired components are maintained.

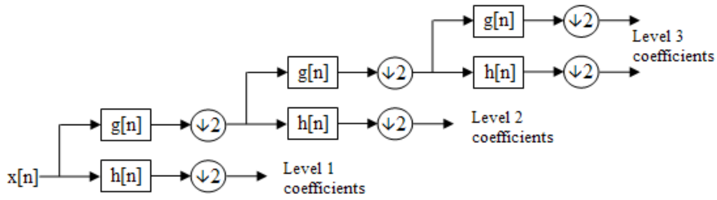

Figure 7: Example of filter bank with N wavelets applied to function decomposition (Extracted from Wikipedia [1])

Wavelet Networks derive logically from this structure and go beyond, offering the possibility to be applied for learning or classification. They present interesting characteristics that will be detailed in the following section and which make them especially relevant for applications in computer vision – such as face recognition or tracking – biology etc.

2.2 Wavelet Networks

Wavelet Networks attempt to combine the properties of the Wavelet decomposition previously described, along with the characteristics of neural networks [10]. Their structure relies on the aforementioned principles –underlying non-linear function approximation– and is given by the equation

$$\tilde{f}(x) = \sum_i w_i \psi_{\mathbf{n}_i}(x) \tag{2}$$

in which the weights w_i represent the coefficients of the network (Fig 8). These are to be tuned as the network learns, in order to give preference to relevant components among the set of N wavelet functions $\mathbf{\Psi} = (\psi_{\mathbf{n}_1}, \psi_{\mathbf{n}_2} ... \psi_{\mathbf{n}_N})$, whereas non-relevant ones are to be penalised.
In this notation, the vector $\mathbf{n}_i$ for each wavelet gathers its corresponding parameters, i.e. $\mathbf{n} = (s, u)$ in the case of the 1D decomposition of functions in $\mathbf{L}^2(\mathbb{R})$, or $\mathbf{n} = (s_x, s_y, u_x, u_y, \theta)$[1] for 2D images in $\mathbf{L}^2(\mathbb{R}^2)$.

Definition: At the sight of equation (2), the wavelet network is completely defined by the tuple $(\mathbf{\Psi}, \mathbf{w})$. Its optimised components may be obtained by calculating the weights w_i and wavelet parameters $\mathbf{n}$ that minimise the least-square error function for $f(\mathbf{x})$, i.e. the ones that make the model fit better to the original function f:

$$\min_{\mathbf{w},\mathbf{n}} \parallel f(\mathbf{x}) - \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i}(\mathbf{x}) \parallel^2$$

This calculation hence implies finding the most suitable N wavelets on which to project, along with the weight that each component ought to be given –how much it ought to contribute to the overall description of $f(\mathbf{x})$– in order to maximise the approximation $\tilde{f}(x)$. Such procedure is to be carried out during the *learning* phase so as to adapt the network to the set of training data points $\{T_n\}$.

Let us remind that minimising the least-square error function in $\mathbf{L}^2(\mathbb{R}^2)$ corresponds to finding the function $\tilde{f}(\mathbf{x})$ that reduces the euclidean distance to each point in the training data, i.e. the one that minimises the lost **energy** that is due to the approximation.

[1] θ represents the orientation in the 2D plane

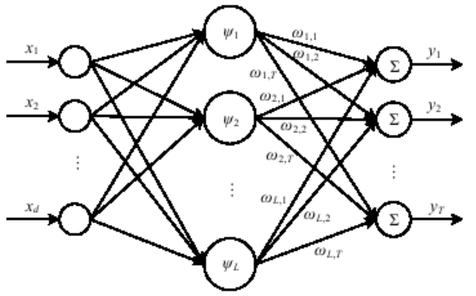

Figure 8: Wavelet Network structure (similar to the perceptron). Example with multiple inputs and multiple outputs (Extracted from Chen 2006 [2])

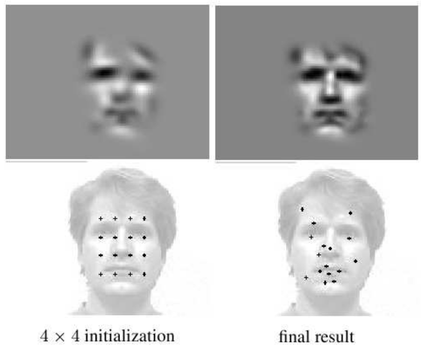

Figure 9: (Left) Initial distribution of the wavelets (4×4) and (Right) their positions after optimisation (Extracted from V. Kruger 1999 [5])

Weight Distribution: It may be noted from equations (1) and (2) that there is a tight link between the weights of the network w_i and the wavelet decomposition $< f, \psi_{\mathbf{n}_i} >$. As a matter of fact, the values for w_i are automatically provided by the wavelet transform [10].

As a result, the final distribution of weights (after optimisation) inherits the suitability of wavelets for 'feature detection'. The w_i are thus automatically tuned so as to **prioritize projections that highlight transients**, whereas wavelets encoding less relevant parts of images are penalised. The result is a network which is *directly related to the underlying image structure* [5], and in which all parameters $(w_i, \mathbf{n})$ are *jointly* fitted from data.

This characteristic is clearly illustrated in Fig 9 (right image) where the 16 wavelets have been distributed at locations where the biggest transients may be encountered, such as for instance above the eyebrows, at the limit of the hair or around the mouth.

Algorithm: The *initialisation* of the network may consist of an arbitrarily distributed set of wavelets, i.e. homogeneously distributed test functions with random orientations θ and constant dilations and scaling factors. Nevertheless, since wavelets are rapidly-vanishing functions, the parameters may need to be constrained in order to prevent degenerated wavelet shapes [3].

As for the **updating** requirements, *backpropagation* algorithms are commonly employed to tune the parameters. This technique is broadly used for teaching feed-forward neural networks, and is thus perfectly adapted to our wavelet-based case. In addition, it presents many advantages such such as quick convergence time. Likewise, it may be possible to make use of optimisation algorithms such as the *Levenberg-Marquardt* method, which is especially suitable for least-square problems.

3 Applications

The properties that have been emphasised until now make Wavelet Networks especially interesting for a wide range of applications in engineering, computer science or biology [9]. These may range from *classification*, to *feature extraction* or *approximation* of complex non-linear functions. Some relevant examples will be further discussed in the following lines, in order to give a flavour of the possibilities offered by this technique.

3.1 Computer Vision

Many probabilistic approaches have been developed for computer vision, such as neural networks, PCA or Eigenfaces. These methods tend to learn the variance of grey-value pixels over a set of training data, and then make use of that pixel-basis knowledge to classify new images. This is performed independently of the object itself.

The interest of Wavelet Networks relies on their ability to be **directly related to the underlying structure of the image** [5]. The wavelets functions – on which the model is built – are 'natural' feature detectors [7] and are independent, for instance, of illumination changes. Figure 10 points out how wavelets do, for instance, automatically adopt the *orientation* θ *of the image*. Moreover, they provide a resolution that may be tuned in order to concentrate on given regions, making them specially suitable for surveillance or tracking applications.

An example for face tracking is described by Krueger et.al. [5] in which 16 sets (4×4) of 4 Gabor wavelets (with corresponding orientations $0, \pi/4, \pi/2$ and $3\pi/4$) are distributed within the region of interest (the *face*). The resulting projections are then calculated using a neural network and the solutions of each filter response are obtained. In order to be able to track the face, an update is performed for each frame, which implies (1) a re-parametrisation of the model (so as to follow the movement of the target) and (2) an optimisation of the previous weights to account for changes in the image. This leads to the tracking shown in figure 11.

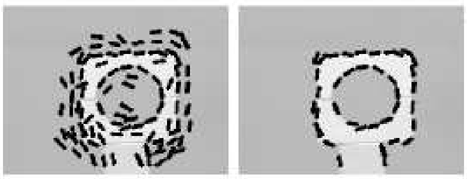

Figure 10: Different wavelets sharing the orientation of the feature in the image (left). Best ones automatically selected (right) (extracted from Krueger 2002 [5])

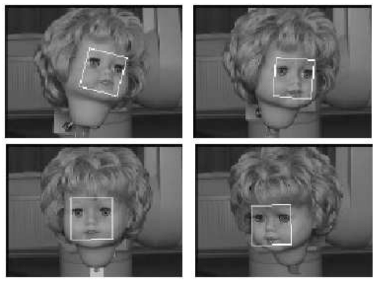

Figure 11: Examples of Face tracking (extracted from Krueger 2002 [5])

3.2 Engineering

Robot control: Robot motion is described by complex non-linear dynamics equations which include time-dependent parameters and system uncertainties, as observed in the differential description:

$$u = M(q)\ddot{q} + C(q, \dot{q})\dot{q} + g(q)$$

In this framework, *approximations* through non-linear networks turn out to be very useful for learning control patterns [4], solving inverse kinematics problems [2] and synthesising correct behavior. This purpose has been previously studied and solved by means of neural networks based on radial basis functions (RBF), i.e. functions that depend only on the distance to a reference point (or centre) $\mathbf{c}_i$:

$$\tilde{f}(\mathbf{x}) = \sum_{i=1}^{N} w_i \phi(\| \mathbf{x} - \mathbf{c}_i \|)$$

However, for a given function, the RBF network may not be unique, nor particularly efficient. The model developed by Katic et.al. [4], for example, replaces this activation function by a wavelet-based network which then plays the role of a **robust controller**, helps compensate uncertainties when the system is in contact with the environment, and yields much more computational efficient results. Promising results have been achieved in this context for controlling manipulator robots such as the one in figure 12.

Product quality monitorisation: Thomas et.al. [8] extends the concept of wavelet networks to cope with the requirements of production lines. This purpose entails monitoring large number of non-stationary signals which are obtained form sensors, and performing **feature extraction** and **classification** so as to come up with a *diagnosis* system.

In this environment, the wavelet network concentrates on transients and is used to extract specific features from the signals and to *recognise* the state of the system. This model makes use of different wavelet networks, one for each feature group, as illustrated in Figure 13, and includes a geometry-based criterion for model selection. Successful results have been achieved for engine knock detection. This asserts, from a more general perspective, the interest of wavelet networks for dealing with industrial problems which involve noisy sensors and non-stationary readings.

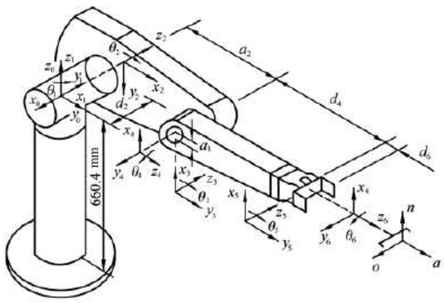

Figure 12: Manipulator PUMA560 for which Inverse kinematics may be calculated through wavelet networks (Extracted from Chen 2006 [2])

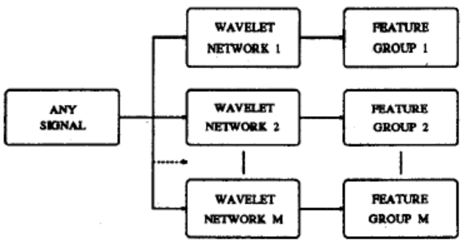

Figure 13: Feature extraction phase with several wavelet networks (Extracted from Thomas 1996 [8])

4 Conclusion

This review has presented the Wavelet transform and its use within feed-forward neural networks. The mathematical foundations have emphasized important properties regarding feature detection and non-lineal function approximation, which have been shown to be of great interest for applications in engineering, robotics and computer vision. Examples of face tracking or dynamic control have been highlighted as proof of what this technique may lead to in the future.

Still, as stated in [6], '*the world of transients is considerably larger and more complex than the garden of stationary signals*', and this review does certainly not attempt to give a full overview of all the techniques currently available in this field. It hopes, nevertheless, to have provided a clear background of the difference with other approaches, and to motivate the reader to continue deepening in the possibilities offered by this domain.

References

[1] Bank filters, 2009. http://en.wikipedia.org/wiki/Wavelet (last modified: January 30, 2009).

[2] H. Cheng, W. shan Chen, and T. Xie. Wavelet network solution for the inverse kinematics problem in robotics manipulator. *Journal of Zhejiang University SCIENCE A*, 4(7):525–529, 2006.

[3] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the optical society of America*, pages 1160–1169, 1985.

[4] D. Katic and M. Vukobratovic. Wavelet neural network approach for control of non-contact and contact robotic tasks. *Intelligent Control, 1997. Proceedings of the 1997 IEEE International Symposium on*, pages 245–250, Jul 1997.

[5] V. Krueger and G. Sommer. Gabor wavelet networks for efficient head pose estimation. *Image and Vision computing*, 20:665–672, 2002.

[6] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[7] B. Manjunath and rama Challapa. A unified approach to boundary perception: Edges, textures and illusory contours. In *IEEE Transactions on Neural Networks*, volume 1, pages 96–108, Jan 1993.

[8] J.-H. Thomas and B. Dubuisson. A diagnostic method using wavelets networks application to engine knock detection. *Systems, Man, and Cybernetics, 1996., IEEE International Conference on*, 1:244–249 vol.1, Oct 1996.

[9] M. Thuillard. A review of wavelet networks, wavenets, fuzzy wavenets and their applications. *Advances in Computational Intelligence and Learning*, pages 43–60, 2002.

[10] Q. Zhang and A. Benveniste. Wavelet networks. In *IEEE Transactions on neural networks*, volume 3, pages 889–898, 1992.