Signal Processing 136 (2017) 2–15

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

The development of the quaternion wavelet transform

P. Fletcher*, S.J. Sangwine

School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

ARTICLE INFO

Keywords:
Quaternion wavelet transform
Quaternion STFT

ABSTRACT

The purpose of this article is to review what has been written on what other authors have called *quaternion wavelet transforms* (QWTs): there is no consensus about what these should look like and what their properties should be. We briefly explain what real continuous and discrete wavelet transforms and multiresolution analysis are and why complex wavelet transforms were introduced; we then go on to detail published approaches to QWTs and to analyse them. We conclude with our own analysis of what it is that should define a QWT as being truly quaternionic and why all but a few of the “QWTs” we have described do not fit our definition.

1. Introduction

In this article we try to show how quaternion wavelet transforms (QWTs) have been developed. Wavelet transforms represent signals using a linear combination of basis functions called *wavelets*, whose principal characteristic is that they are localised in time or space. Unlike a representation using periodic basis functions such as sines and cosines, wavelet transforms allow localised signal content to be analysed.

As we note at the beginning of Section 4, most of the theory about wavelets and wavelet transforms, including that about complex wavelet transforms ($\mathbb{C}$WTs) and QWTs, has been developed for real-valued signals and greyscale images.

However, we are concerned with signals and images which require more than one component per sample or pixel, for example quaternion signals and, in particular, colour images represented as arrays of quaternions: QWTs for processing these will have different properties from the majority of the QWTs that have appeared so far. As well as discussing QWTs for quaternion signals and arrays, we also briefly consider $\mathbb{C}$WTs for complex signals.

Throughout the paper, when we refer to signals and time, it should also be understood that we may mean images and space (in the sense of position within an image). Most of what we cover generalises to the 2-D case of images.

In Section 2 we briefly cover the history of real wavelets and describe their properties. There are three types of wavelet analysis, using continuous wavelets, discrete wavelets and multiresolution analysis, and we discuss each of these in turn. There are some problems with real wavelets and in Section 3 we see how complex wavelets were introduced to solve them. In Section 4 we survey all articles that have contributed to the development of QWTs and describe what each article's authors have done. There is no single approach to QWTs and in Section 5.4 we say what we believe a “true” QWT should and should not look like. In Section 6 we present our conclusions.

2. Classical ($\mathbb{R}$-valued) wavelet transforms

We start by reviewing briefly the main ideas of wavelets and wavelet transforms in the real-valued case, in order to provide some context for the rest of the paper. A much fuller treatment is given by, e.g., Kovačević et al. [59].

2.1. Background

The Fourier transform (FT) gives information about the frequency content of a signal, but nothing about where in time or space its different constituent frequencies occur. In some applications, for example with non-stationary signals where some frequency content is present only for a limited time, it would be desirable to know the distribution of frequencies over time or space, which led to the introduction of the short-time Fourier transform (STFT) by Gabor [29]. A “sliding window” is introduced into the FT, initially centred at time 0, say; the signal is assumed to be approximately stationary in the window and its FT is found. The window is then shifted by t and the FT of the new section of signal is found, and so on until the whole signal has been covered. In continuous time and frequency the STFT in 1-D, centred on time t, can be expressed as

$$\mathcal{F}_{\text{STFT}} f(\omega, t) = \int_{-\infty}^{\infty} f(s) g(s - t) \mathrm{e}^{-j\omega s} \mathrm{d}s, \tag{1}$$

where $g(\cdot)$ is the window function. In discrete time the STFT becomes

* Corresponding author.
E-mail addresses: pfletc@essex.ac.uk (P. Fletcher), sjs@essex.ac.uk (S.J. Sangwine).

http://dx.doi.org/10.1016/j.sigpro.2016.12.025
Received 22 March 2016; Received in revised form 8 December 2016; Accepted 26 December 2016
Available online 29 December 2016
0165-1684/

$$F_{\text{STFT}}(\omega, t) = \sum_{m=-\infty}^{\infty} f[m]g[m-t]\mathrm{e}^{-j\omega m}. \tag{2}$$

The function $g(\cdot)$ or $g[\cdot]$ will always be even in practice and some authors would write $g(t-s)$ and $g[t-m]$ above. Gabor experimented with a number of different functions for the window $g(\cdot)$ and found the best he could do was to use a Gaussian; the STFT with this window function is now called the Gabor transform. We cannot know the exact frequency at a given time and Heisenberg's Uncertainty Principle applies per Gabor [30]:

$$\Delta t \Delta\omega \geq \frac{1}{2}, \tag{3}$$

where Δt is the uncertainty in time and $\Delta\omega$ is the uncertainty in angular frequency. The Gabor transform is optimal in the sense that this inequality theoretically becomes an equality when $g(\cdot)$ is a Gaussian.

A major drawback of the STFT is that once the window is chosen, its resolution is fixed. There are two extremes to consider: a high frequency signal with a period less than the width of the window and only a few oscillations would have a relatively large uncertainty as to its actual position; and a lower frequency signal with a period longer than the width of the window would not actually be detected at all. Heuristically, the (discrete) wavelet transform is similar to an STFT, but one with a range of different window sizes: a larger number of short windows to capture the detail at higher frequencies and a smaller number of long windows for the lower frequencies.

The term "wavelet" had already been used for many years by geophysicists, e.g., Ricker [88], to refer to a single component of a seismogram when in the early 1980s, e.g., Grossmann and Morlet [44], the mathematics of wavelets was developed to allow them to be used as a tool in signal processing. The simplest wavelet is the *Haar wavelet*, which appears to have been so-named in the 1970s or 1980s: Haar [46] studied systems of orthogonal functions using a set of orthogonal rectangular basis functions, with each basis function consisting of a short positive pulse followed immediately by a short negative pulse:

$$\psi(t) := \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Later, after the development of wavelets, this Haar wavelet was generalised to $\psi_{\alpha,\beta}(t) := 2^{-\alpha/2}\psi(2^{-\alpha}t - \beta)$, where $\alpha, \beta \in \{0\} \cup \mathbb{Z}^+$.[1] The scale factor $2^{-\alpha/2}$ ensures that $\int_{-\infty}^{\infty} |\psi_{\alpha,\beta}(t)|^2 \mathrm{d}t = 1$. $\psi(t)$ is called the *mother wavelet* and the $\psi_{\alpha,\beta}(t)$ functions, *daughter wavelets*; α is a scale parameter and β is a translation parameter. In Fig. 1, we illustrate the Haar mother wavelet with two examples of daughter wavelets and use this notation.

There was little interest in Haar's rectangular pulse until it was picked up by Lévy [65] as an improvement on the Fourier basis functions for studying the fine detail of Brownian motion: as demonstrated by Pinsky [86], Brownian motion can be expressed as a sum of Haar wavelets.

A square pulse is not the only wavelet that can be used. Strömberg [100] started the development of discrete wavelets beyond what Haar had done and Daubechies [23] introduced families of orthogonal wavelets with compact support. The continuous wavelet transform first appeared in Zweig et al. [124], although that in Goupillaud et al. [42] is the oldest which would be recognised today as a wavelet transform.

[1] Some authors write $\psi_{\alpha,\beta}(t) = 2^{\alpha/2}\psi(2^{\alpha}t - \beta)$, but the notation we have used is as per Daubechies [24, p. 10].

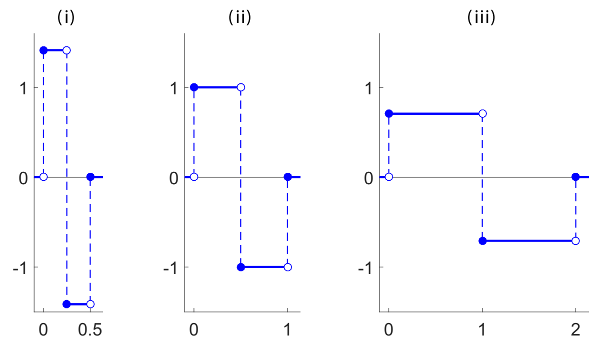

Fig. 1. The Haar mother wavelet (ii) and two levels of Haar daughter wavelets: (i) $\psi_{-1,0}(t) = \sqrt{2}\psi(2t)$, (ii), $\psi_{0,0}(t) = \psi(t)$, (iii) $\psi_{1,0}(t) = \frac{1}{\sqrt{2}}\psi\left(\frac{1}{2}t\right)$.

2.2. Properties of wavelets

Many wavelets have been developed, each with properties suited to particular applications. All have the property of *localisation* in space and time and some have infinite support, but the most popular, as mathematical objects if not for applications according to Blatter [10, p. 6], have *finite support*. A wavelet having finite support simply means that there is a finite interval, outside of which its amplitude is zero. A wavelet transform should be able to analyse a signal (or image) at different scales and so the underlying wavelets need to be localised in spatial frequency. It also needs to encode where in time (or space) these frequencies occur and so the wavelets need to be localised in time (or space) as well. A 1-D example of space and time localisation would be a music score, which shows which musical notes and hence sound frequencies need to occur and at what times. Wavelet analysis of the music would theoretically allow the music score to be reconstructed but Fourier analysis of the same music would not, since it would not reveal the locations of different frequency content due to individual notes.

The wavelet functions are chosen from $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, the space of measurable functions that are absolutely and square integrable:

$$\int_{-\infty}^{\infty} |\psi(t)| \mathrm{d}t < \infty, \quad \text{and} \quad \int_{-\infty}^{\infty} |\psi(\mathrm{t})|^2 \mathrm{d}t < \infty.$$

In addition, a wavelet must have zero mean and a squared norm of unity, so:

$$\int_{-\infty}^{\infty} \psi(t) \mathrm{d}t = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} |\psi(\mathrm{t})|^2 \mathrm{d}t = 1.$$

We define a wavelet as

$$\psi_{a,b}\colon \mathbb{R} \to \mathbb{C}, \quad t \mapsto \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right),$$

where $(a, b) \in \mathbb{R}^+ \times \mathbb{R}$ [10, p. 14]. Note that this definition of $\psi_{a,b}$ is slightly different from the one we used for the $\psi_{\alpha,\beta}$ of the Haar wavelet above: if $\psi_{a,b}$ were that Haar wavelet, we would have $a = 2^{\alpha}$ and $b = 2^{\alpha}\beta$. Both definitions are in use and we also call a and b the *scale* and *translation* parameters respectively. Fig. 1

2.3. The continuous wavelet transform

The continuous wavelet transform

$$\mathcal{W}_{\psi}f\colon \mathbb{R}^+ \times \mathbb{R} \to \mathbb{C}, \quad (a, b) \mapsto \mathcal{W}_{\psi}f(a, b)$$

of a signal f is defined as

$$\mathcal{W}_{\psi}f(a, b) := \langle f, \psi_{a,b}\rangle = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}\mathrm{d}t, \tag{4}$$

where the variables are as we defined in Section 2.2. The result is a data array