The Berkeley Wavelet Transform: A biologically-inspired orthogonal wavelet transform

Ben Willmore[1,4], **Ryan J Prenger**[2], **Michael C-K Wu**[3], and **Jack L Gallant**[1]

[1]Department of Psychology & Helen Wills Neuroscience Institute, University of California, Berkeley, 3210 Tolman Hall #1650, Berkeley, California 94720-1650, USA

[2]Department of Physics, University of California, Berkeley, 3210 Tolman Hall #1650, Berkeley, California 94720-1650, USA

[3]Biophysics Graduate Group, University of California, Berkeley, 3210 Tolman Hall #1650, Berkeley, California 94720-1650, USA

Abstract

We describe the Berkeley Wavelet Transform (BWT), a two-dimensional triadic wavelet transform. The BWT comprises four pairs of mother wavelets, at four orientations. Within each pair, one wavelet has odd symmetry, and the other has even symmetry. By translation and scaling of the whole set (plus a single DC term), the wavelets form a complete, orthonormal basis in two dimensions.

The BWT shares many characteristics with the receptive fields of neurons in mammalian primary visual cortex (V1). Like these receptive fields, BWT wavelets are localised in space, tuned in spatial frequency and orientation, and form a set that is approximately scale invariant. The wavelets also have spatial-frequency and orientation bandwidths which are comparable with biological values.

Although the classical Gabor wavelet model is a more accurate description of the receptive fields of individual V1 neurons, the BWT has some interesting advantages. It is a complete, orthonormal basis, and is therefore inexpensive to compute, manipulate and invert. These properties make the BWT useful in situations where computational power or experimental data are limited, such as estimation of the spatiotemporal receptive fields of neurons.

Keywords

simple cell; complex cell; triadic; quadrature

1 Introduction

Wavelets are basis functions that are localized in space and spatial frequency. They can be used to analyse the spatial-frequency content of signals without discarding position

[4]Present address: Department of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road, Oxford OX1 3PT, United Kingdom. benjamin.willmore@dpag.ox.ac.uk

information, and so they provide a useful compromise between representation in 'pixel space' and Fourier space. Studies of natural image statistics have shown that wavelets are particularly appropriate for encoding natural images, because wavelet-like features correspond to statistically independent, 'sparse' structure in natural images (Field, 1987; Olshausen & Field, 1996; Bell & Sejnowski, 1997).

Primary visual cortex (V1) of the mammalian visual system contains neurons which represent the structure of the retinal image in terms of a wavelet basis. Gabor wavelets (Gabor, 1946; Daugman, 1980; Marcelja, 1980) provide an accurate description of the behavior of neurons in V1 (Jones & Palmer, 1987; Smyth et al., 2003). For computationally-intensive purposes however, Gabor wavelets have a significant drawback: standard Gabor pyramids are overcomplete and therefore linearly dependent. Thus, to completely represent an image containing n pixels, a larger number of Gabor wavelets is required, typically around $5n$ (Navarro et al., 1996).

We aimed to estimate the nonlinear Spatio-Temporal Receptive Fields (STRFs) of neurons in cortical areas V2 and V4 (Willmore et al., 2005, 2006). These neurons receive their input (directly and indirectly) from neurons in V1, and so it is natural to describe them in terms of the STRFs of V1 neurons. The Gabor transform is the obvious choice, but in practice we find that the overcompleteness of the Gabor transform is a disadvantage for this purpose. A STRF model that incorporates the Gabor transform will contain a relatively large number of coefficients compared to a similar model that uses an orthogonal basis. When estimating the values of these coefficients based on a limited amount of neurophysiology data, this means that each estimated coefficient will be relatively inaccurate. To estimate the best possible models using limited data, it is more efficient to use a complete, orthogonal basis.

We therefore developed a wavelet basis which is complete and orthogonal (thus making efficient use of neurophysiology data), and maintains as many of the characteristics of V1 STRFs as possible. We describe here the resulting transform — the Berkeley Wavelet Transform or BWT.

2 Modeling the wavelet code in V1

To develop a wavelet basis that has similar properties to the neural code in visual area V1, we first need to discuss the neural code itself. In this section, we summarize the characteristics of V1 neurons and define a set of criteria that should be met by a computational model of these neurons.

2.1 Simple cells

The responses of cortical simple cells result from approximately linear summation of the luminance of the retinal image (Movshon et al., 1978b) (ignoring adaptation effects). Their tuning can be described by an LN model where the linear filter is a two-dimensional generalization of the Gabor filter (Daugman, 1980; Marcelja, 1980; Jones & Palmer, 1987):

$$\mathbf{g}_{x_0,y_0,\lambda,\theta,\phi,\sigma}(x,y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x'+y')^2}{2\sigma^2}\right]\sin(x'/2\pi\lambda-\phi)$$

(1)

where $x' = (x-x_0)\cos\theta$ and $y' = (y-y_0)\sin\theta$ are coordinate axes rotated to an angle θ.

Since neurons produce only positive responses, a model of the simple cell response, r, to an image, $\mathbf{s}$, is:

$$r = |\mathbf{g}.\mathbf{s}|^+$$

(2)

This 2D generalization of the Gabor filter (Figure 1A) expresses the primary tuning characteristics of individual simple cells. The filter is *spatially localized*: it has a location (x_0, y_0), and is bounded by a Gaussian window (spread σ) centered on that location. The filter has *bandpass spatial-frequency tuning* and *bandpass orientation tuning*. The sinusoidal modulation determines the preferred spatial frequency, $1/\lambda$, and orientation, θ. The spread, σ, of the Gaussian window determines the bandwidth. Finally, the filter has a preferred phase that is determined by the phase, ϕ, of the sinusoidal modulation (Figure 1E).

2.2 Complex cells

Complex cells share most of the characteristics of simple cells. Unlike simple cells, however, complex cell responses are approximately invariant to the spatial phase of visual stimuli (Movshon et al., 1978a). Complex-cell responses can be modeled as the sum of the squares of the responses of a self-similar pair of 2D Gabor filters (Pollen & Ronner, 1981, 1983; Watson & Ahumada, 1983; Adelson & Bergen, 1985) in quadrature phase (separated by $\pi/2$ radians):

$$r = (\mathbf{g}_{\phi=0}.\mathbf{s})^2 + (\mathbf{g}_{\phi=\pi/2}.\mathbf{s})^2$$

(3)

Thus, a wavelet transform that can effectively model phase-invariance in complex cells should contain pairs of wavelets in *quadrature phase*.

2.3 Characteristics of the population

Equations 2 and 3 specify a functional form for models of the spatial receptive fields of simple and complex cells. We can specify the parameters of these functions by observing the variation of SRFs over the V1 population. First, the position (x, y), wavelength λ, orientation preference θ and (for simple cells) phase preference ϕ are all distributed continuously, approximately equiprobably, and approximately independently across the available ranges. The spread, σ, of biological SRFs is approximately proportional to the wavelength, λ (although there is significant individual variation between cells). As a result, low-frequency SRFs are qualitatively similar to high-frequency SRFs; i.e. the representation

is *rotationally similar*. Orientation is independent of the other parameters, so the representation is also *similar across orientation*.

Studies of neurons in macaque V1 have determined the mean relationship between the spread, σ, and the mean wavelength, λ. The mean bandwidth of spatial-frequency tuning (full width at half height) is 1.5 octaves (De Valois et al., 1982); the median bandwidth of orientation tuning is 23.5° (Ringach et al., 2002).

2.4 Criteria for modeling V1 cells

An ideal model of the behavior of a V1 population should have the same characteristics as the simple and complex cells described above, i.e.:

- Spatial localization

- Bandpass spatial-frequency tuning, with bandwidth $\approx$ 1.5 octaves

- Bandpass orientation tuning, with bandwidth $\approx$ 23.5°

- Similarity across scale

- Rotational similarity

- Quadrature phase

3 Definition of the BWT

The BWT is a complete, orthonormal wavelet transform which approximates all of the properties described above. It is defined by a set of mother wavelets and a dilation equation that specifies the scaling and translation of the mother wavelets.

3.1 Mother wavelets

The BWT comprises eight mother wavelets, in four pairs. Each pair has a different orientation — 0°, 45°, 90°, and 135° — represented here by $\Theta = \{\theta\} = \{|, /, -, \backslash\}$ respectively. Within each pair, one wavelet is odd-symmetric and one is even-symmetric. We represent these phases by $\Phi = \{\phi\} = \{o, e\}$. The set of eight mother wavelets can be written as $\beta_{\theta,\phi}$. We define a two-dimensional pulse, $u(x, y)$:

$$u(x,y) = \begin{cases} 1, \text{if } 0 < x \leq 1, 0 < y \leq 1 \\ 0, \text{otherwise} \end{cases} \tag{4}$$

Then, the four pairs of mother wavelets, $\beta_{\theta,\phi}$, are:

$$\beta_{|,o}(x,y) = [-u(x,y/3) + u(x-2,y/3)]/\sqrt{6} \tag{5}$$

$$\beta_{|,e}(x,y) = [-u(x,y/3) + 2u(x-1,y/3) - u(x-2,y/3)]/\sqrt{18} \tag{6}$$

$$\beta_{/,o}(x,y) = [-(u(x,y-2)+u(x-1,y)+u(x-2,y-1))+u(x,y-1)+u(x-1,y-2)+u(x-2,y)]/\sqrt{6}$$

(7)

$$\beta_{/,e}(x,y)$$
$$=[-(u(x,(y-1)/2)+u(x-1,y)+u(x-1,y-2)+u(x-2,y/2))+2(u(x,y)+u(x-1,y-1)+u(x-2,y-2))]/\sqrt{18}$$

(8)

$$\beta_{-,o}(x,y) = [u(x/3,y)-u(x/3,y-2)]/\sqrt{6}$$ (9)

$$\beta_{-,e}(x,y) = [-u(x/3,y)+2u(x/3,y-1)-u(x/3,y-2)]/\sqrt{18}$$ (10)

$$\beta_{\backslash,o}(x,y) = [-(u(x,y)+u(x-2,y-1)+u(x-1,y-2))+u(x,y-1)+u(x-1,y)+u(x-2,y-2)]/\sqrt{6}$$

(11)

$$\beta_{\backslash,e}(x,y)$$
$$=[-(u(x,y/2)+u(x-1,y)+u(x-1,y-2)+u(x-2,(y-1)/2))+2(u(x,y-2)+u(x-1,y-1)+u(x-2,y))]/\sqrt{18}$$

(12)

Note that $\beta_{-,o}$ and $\beta_{-,e}$ can be obtained by transposing x and y in Equations 5 and 6. Also $\beta_{\backslash,o}$ and $\beta_{\backslash,e}$ can be obtained by a 90° rotation of the coordinate axes about (2, 2) in Equations 7 and 8 (Figure 2A). The wavelets are defined to be zero outside multiples of a 3 × 3 grid, and are therefore compactly supported (Daubechies, 1988). Appendix A demonstrates that they meet the admissibility condition for wavelets.

3.2 Discrete form

The mother wavelets are piecewise constant functions, so each wavelet, $\beta_{\theta,\phi}$, can be completely described by a 3×3 matrix, $b_{\theta,\phi}$:

$$b_{|,o} = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \tag{13}$$

$$b_{|,e} = \frac{1}{\sqrt{18}} \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix} \tag{14}$$

$$b_{/,o} = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{bmatrix} \tag{15}$$

$$b_{/,e} = \frac{1}{18} \begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix} \tag{16}$$

$$b_{-,o} = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{17}$$

$$b_{-,e} = \frac{1}{\sqrt{18}} \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix} \tag{18}$$

$$b_{\backslash,o} = \frac{1}{\sqrt{6}} \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} \tag{19}$$

$$b_{\backslash,e} = \frac{1}{\sqrt{18}} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \qquad (20)$$

3.3 Dilation equation

The mother wavelets, $\beta_{\theta,\phi}$, are scaled and translated using the dilation equation to produce daughter wavelets, $\beta_{\theta,\phi}^{m,n,s}$, at multiple positions (m, n) in the x-y plane, and multiple scales, s:

$$\beta_{\theta,\phi}^{m,n,s} = \frac{1}{s^2} \beta_{\theta,\phi} \left(3^s(x-m), 3^s(y-n) \right) \qquad (21)$$

The BWT uses triadic scaling, i.e. the sizes of the daughter wavelets are scaled by powers of 3. To form an orthogonal set, the possible translations in the x-y plane are locked to integer multiples of the wavelet size, 3^s.

3.4 DC term

All of the BWT wavelets have zero mean, and so a single DC term is required to represent the mean value of an image:

$$\beta_0 = [u(x/3, y/3)] \sqrt{9} \qquad (22)$$

Note that this DC term is not subject to the dilation equation; instead, it is applied only once, at the largest spatial scale.

3.5 Symmetry and orthogonality

Figure 2B shows the 3×3 Discrete Fourier transforms (DFTs) of the BWT mother wavelets. The wavelets are mutually orthogonal, and localized in spatial-frequency and orientation. Additionally, the DFTs of $b_{|,o}$, $b_{/,o}$, $b_{-,o}$ and $b_{\backslash,o}$ are imaginary and odd, indicating that these wavelets are real and odd. The DFTs of $b_{|,e}$, $b_{/,e}$, $b_{-,e}$ and $b_{\backslash,e}$ are real and even, indicating that these wavelets are also real and even.

We can use these symmetry properties to demonstrate the orthogonality and completeness of the BWT basis (Appendix B). Since the BWT forms a complete, orthogonal set, it is also a tight frame.

3.6 Reconstruction

Since the BWT is a complete, orthonormal set, it is self-inverting. An image can be reconstructed from its BWT coefficients using:

$$I(x,y) = \sum_{\theta \in \Theta, \phi \in \Phi} \sum_{m,n,s=0}^{\infty} w_{m,n,s}^{\theta,\phi} \beta_{\theta,\phi} \left(3^s (x-m), 3^s (y-n)\right) \tag{23}$$

where $w_{m,n,s}^{\theta,\phi}$ are the BWT coefficients representing the image.

4 Characteristics of the BWT

The neural code in V1 is a non-orthogonal, overcomplete representation. The BWT attempts to model this code as closely as possible while maintaining completeness and orthogonality. The BWT therefore represents a compromise between biological accuracy and computational efficiency. Many other wavelet transforms exist, some of which have attempted to achieve a similar compromise. Here we discuss the characteristics of the BWT, and compare the properties of the BWT to those of related wavelet transforms.

4.1 Uniqueness

It is interesting to consider whether the BWT is the only transform of its kind, or whether it is one member of a family of similar transforms. Inspection of the 3×3 Discrete Fourier Transforms (DFTs) of the mother wavelets shows that (apart from a sign flip) they are uniquely specified by the criteria that the 3×3 DFT power spectra of the wavelets should be a pair of delta functions and that the wavelets should come in odd-even pairs (see Figure 2B). In other words, the BWT is the only odd-even basis that has perfect spectral localization in this space (though the spectral localization breaks down at higher spectral resolution, as shown in Figure 4B).

The wavelets can also be uniquely specified by fairly simple criteria in pixel space. The BWT is the only orthonormal set of 3×3 wavelets in which each wavelet (apart from the D.C.) varies along only one axis (0°, 45°, 90° or 135°) and is either odd- or even-symmetric about that axis. The BWT coefficients are derived (and their uniqueness demonstrated) in Appendix C.

A family of related transforms could be constructed by linear combination of the members of each odd-even pair of BWT basis functions. These would have similar properties to the BWT, but the wavelets would not have odd and even symmetry. It may also be possible to construct transforms similar to the BWT on a 4×4 or 5×5 grid.

4.2 Computational issues

The BWT has two properties that make it particularly inexpensive to compute, relative to other wavelet transforms. First, all BWT wavelets are mutually orthogonal, and so arbitrary subsets of the transform can be computed independently. In particular, the wavelets are orthogonal across scale, and so different scales of the transform can be computed separately. Second, the BWT wavelets are piecewise constant. Thus, to calculate the transform at scale *s*, one can downsample the image to the appropriate size and then calculate the scalar products. This is approximately 9^s times less computationally expensive than calculating the

scalar products at full resolution. A disadvantage of being piecewise constant is that that, like Haar wavelets (Haar, 1910), the BWT pyramid requires a relatively large number of coefficients for approximating smooth functions.

The BWT is a triadic (3-fold) wavelet pyramid, whereas most wavelet pyramids are dyadic (though for another example of a triadic pyramid, see Rao et al. 1998). The primary advantage of the triadic structure is that it decomposes images into locally even-symmetric and odd-symmetric structure. A disadvantage of the triadic structure is that the BWT operates on images whose side lengths are powers of 3, and so images with side lengths that are powers of 2 need to be resampled before use.

4.3 Rotational similarity

Most wavelets are defined in one dimension, and therefore require generalization to tile the two-dimensional plane. Previous 2D wavelet pyramids have usually taken one of three approaches to this generalization.

One approach is to take a one-dimensional wavelet pyramid in x and y and combine all pairs of wavelets separably (i.e. by taking the product of a 1D wavelet in x and a 1D wavelet in y). This produces a complete, orthogonal pyramid in two dimensions, but tends to violate our criteria of similarity across scale and orientation. For example, although the 1D Haar transform resembles cortical SRFs, the separable 2D generalization of the Haar transform (Figure 3A) contains basis functions with different shapes. These violate the similarity criteria, and do not generally resemble the cortical code.

Since the separable approach does not produce a biologically relevant wavelet code, models of V1 function (Watson, 1987; Porat & Zeevi, 1988) have taken a different approach: they take a Gabor-like mother wavelet and rotate it in the x-y plane. This produces a rotationally homogeneous representation that meets our criteria of similarity across orientation and scale. However, since arbitrary rotations of the mother wavelet are not generally orthogonal to one another, the resulting two-dimensional transform will not generally be orthogonal.

A third approach (Adelson et al., 1987) is to separably combine the mother wavelets before translating and scaling them. A one-dimensional mother wavelet, $\Psi(x)$ can be generalized to 2D by applying a windowing function along the orthogonal dimension, $\Xi(y)$. A complete 2D transform can then be created by translating and scaling mother wavelets $\Psi(x).\Xi(y)$, $\Xi(x).\Psi(y)$, and $\Psi(x).\Psi(y)$. Applying this approach to the Haar transform gives what we will refer to as the Two-Dimensional Haar Transform (2DHT; Figure 3B). The axially-aligned wavelets are similar (though not identical) to the oblique wavelets. This generalization of the Haar transform resembles the BWT, and meets many of the criteria specified in Section 2.4.

4.4 Quadrature phase

A disadvantage of the 2DHT is that it contains only odd-symmetric axially-aligned wavelets and even-symmetric oblique wavelets. This leads to awkward representations of even-symmetric axial structure (such as straight lines) and odd-symmetric oblique structure (such as edges).

The lack of even-odd pairs of basis functions also makes the 2DHT awkward for modeling the V1 representation. Simple cells have different phase preferences, representing images in terms of even-symmetric, odd-symmetric and intermediate structure. The 2DHT does not reflect this variety. Moreover, complex cells have phase invariant responses that are usually modeled as the sum of squares of a pair of Gabor wavelets in quadrature phase (Equation 3). There is no way to construct quadrature-phase pair of 2DHT wavelets. Nor is there any other simple way to construct phase invariant responses in terms of the 2DHT. This is a severe disadvantage for modeling complex cells.

The BWT (Figure 3C) is constructed in a similar way to the 2DHT. Having constructed odd and even phase 1D wavelets, $\beta_o(x)$ and $\beta_e(x)$, the 2D transform is produced by adding to the set $\beta_o(y)$, $\beta_e(y)$, and four new oblique wavelets. The oblique wavelets are not a separable combination of the axially-aligned wavelets. Instead, they are chosen to be odd-even pairs that are similar to, and orthogonal to the axially-aligned wavelets. Although the spatial-frequency tuning of the resulting basis functions are not precisely the same, they are qualitatively similar, and form an orthogonal set.

Taken in odd-even pairs (whose position, orientation and scale are identical), the sum of the squares of the mother wavelets, $\beta_o^2 + \beta_e^2$, has no local structure within the wavelet envelope. Thus, odd-even pairs can be combined to produce phase invariant responses by squaring and summing their responses, as in Equation 3. The summed responses depend on the position, scale and orientation of the wavelets, but not on spatial phase within the envelope of the wavelets. This is an approximation to the classic energy model of cortical complex cells (Adelson & Bergen, 1985).

4.5 Frequency tuning

The BWT tiles spatial frequency (Figure 4) in a manner that is qualitatively similar to the Gabor pyramids of Watson (1987) and Porat & Zeevi (1988). However, some compromises have been made to maintain orthogonality. The contours are not circular, indicating that the wavelets are not as well localized in spatial-frequency and orientation as Gabor filters. At lower contour values, the oblique BWT wavelets have nonnegligible contributions from orthogonal orientations. Also, the wavelets within each level of the pyramid do not have precisely the same preferred spatial frequencies. Most importantly, the spatial-frequency preference of the odd-symmetric axially-aligned wavelets is half the preferred spatial-frequency of their even-symmetric counterparts.

The spatial-frequency bandwidths (full width at half height, calculated from the DFT of the wavelets) of the BWT wavelets are 2.2 octaves ($b_{|,o}$, $b_{-,o}$), 1.5 octaves ($b_{|,e}$, $b_{-,e}$), 1.2 octaves ($b_{/,o}$, $b_{\backslash,o}$), and 1.8 octaves ($b_{/,e}$, $b_{\backslash,e}$). These are similar to the band-widths of V1 neurons; the mean V1 bandwidth is 1.5 octaves (De Valois et al., 1982). The orientation bandwidths (full width at $1/\sqrt{2}$ height) are 72° ($b_{|,o}$, $b_{-,o}$), 37° ($b_{|,e}$, $b_{-,e}$), 34° ($b_{/,o}$, $b_{\backslash,o}$), and 48° ($b_{/,e}$, $b_{\backslash,e}$). These are somewhat higher than the values for V1 neurons; the median V1 bandwidth is 23.5° (Ringach et al., 2002). However, the biological distribution is skewed towards higher bandwidths, up to a maximum of about 80 °. All of the BWT wavelets fall within this range.

4.6 Comparison with other orthogonal V1-like transforms

Two other studies have explicitly attempted to produce develop orthogonal transforms that resemble cortical receptive fields. Adelson et al. (1987) investigated the use of transforms based on Quadrature Mirror Filters. The transforms proposed by (Adelson et al., 1987) are localized in space and contain filters with more precise spatial-frequency and orientation tuning than the BWT wavelets. The most promising of these transforms for cortical modeling contains a low-pass filter and even-symmetric high-pass filters at three orientations (0°, 60° and 120°). However, they have rather low spatial-frequency bandwidths compared to the receptive fields of cortical neurons. Also, the filters do not form quadrature pairs, and are not precisely orthogonal. The biggest disadvantage of these filters is that they are defined on a hexagonal sampling grid. To analyze standard images with a square grid, it is necessary to resample them to the hexagonal grid, requiring extra computation and potentially introducing aliasing artifacts.

Another cortex-like orthogonal transform is the Hexagonal Orthogonal-oriented Pyramid (HOP) developed by Watson & Ahumada (1989). The constraints chosen by Watson & Ahumada are very similar to those that determine the structure of the BWT — in particular, they specified that their transform should contain filters in odd-even pairs. These specifications resulted in two transforms (differing in their even-symmetric filters). Both contain oriented high-pass filters at three orientations (0°, 60° and 120°); one of the two contains quadrature pairs. The filters are spatially localized, and better localized in frequency space than the BWT wavelets. The main disadvantage of the HOP is that, like the QMF filters proposed by Adelson et al., the pyramid is based on a hexagonal sampling grid.

4.7 Comparison with Gabor pyramids

The Gabor filter is the classical model of image processing in V1 because Gabor filters possess all of the characteristics of cortical cells specified above. With appropriate parameters, Gabor filters provide excellent models of single V1 neurons (Jones & Palmer, 1987). Several authors have produced biologically-inspired image coding pyramids based on the Gabor filter (Watson, 1987; Porat & Zeevi, 1988; Navarro et al., 1996). These pyramids provide more accurate models of the characteristics of the V1 population than the BWT. Additionally, Gabors have multiple parameters that can be adjusted to precisely model individual cells. In situations where biological fidelity is the only concern, Gabor filters are superior to the BWT. However, as discussed above, Gabor filters are not generally orthogonal to one another. As a result, classical Gabor transforms that span image space are generally overcomplete, containing more basis functions than pixels.

Not all Gabor transforms are overcomplete. By using alternative modulation and windowing functions, it is possible to produce Gabor transforms that are exactly complete (e.g. Ahmed & Fahmy 1998). However, these transforms do not provide the accurate model of cortical receptive fields provided by the standard Gabor transform. And, although they use the same number of coefficients as the BWT, they are not generally orthogonal or self-inverting, i.e. separate sets of analysis and synthesis functions are required for accurate reconstruction. For some purposes, it is desirable to use a moderately accurate biological model that is also a

complete, orthogonal, self-inverting representation. In such cases, the BWT may be preferable to Gabor transforms.

5 Applications

We have shown that the BWT is an orthogonal basis that provides a good basic model of the tuning of cortical neurons. We developed the BWT for use in STRF estimation. Here we discuss the utility of the BWT for this purpose, and for other potential applications.

5.1 Spatiotemporal Receptive Field estimation

We developed the BWT as a tool for estimating the STRFs of neurons in the mammalian visual cortex. Neurons in higher cortical areas such as MT, V2 and V4 receive input from V1, and the spatial tuning of neurons in higher cortical areas is partially produced by combination of the responses of neurons in lower areas. One approach to modeling the responses of these higher-order neurons is to construct a model of the responses of a population of V1 neurons, and then describe the responses of higher neurons in terms of these modeled responses. The resulting model can be interpreted as a STRF that has been estimated in a biologically-relevant space.

We have built a simple nonlinear STRF model of the responses of neurons in cortical areas V2 and V4 (see Figure 5). We take the BWT decomposition of input images, and half-wave rectify the wavelet responses, taking positive and negative responses separately. This is a simple nonlinear model of the half-wave rectified responses of a population of cortical simple cells. We then describe the responses of neurons in V2 and V4 as weighted linear sums of the responses of the model V1 population. We find that this model provides a good description of the responses of neurons in V2 and V4 to natural stimuli (Willmore et al., 2005, 2006).

For construction of a model such as this, a Gabor pyramid might seem to be the ideal choice. However, standard Gabor pyramids are several times overcomplete. Using an overcomplete basis for STRF estimation presents significant problems. First, neurophysiology data is noisy and limited in quantity. Thus, to accurately fit models, it is important to make maximally efficient use of the available data. Using an over-complete basis produces a model with a relatively large number of coefficients. With a limited data set, the power of the data is spread over a relatively large number of coefficients, resulting in relatively large estimation error for the coefficients. Using an orthogonal basis concentrates the power of the data into a smaller number of coefficients. This reduces the estimation error for each coefficient and results in better fits. Also, STRF estimation problems are computationally expensive, making the extra computational cost of using an overcomplete basis particularly undesirable.

The rectification step is crucial to the success of this STRF model for describing nonlinear responses such as those of complex cells and neurons in higher cortical areas. Without rectification, the STRF model would simply be a linear model, and would therefore provide a poor description of complex cell behavior (Movshon et al., 1978a). However, the rectification doubles the number of coefficients in the STRF model. If we had used a 5×

overcomplete Gabor pyramid, rather than the BWT, the number of coefficients would be even larger. For example, using a 27×27 image and a rectified 5× overcomplete Gabor representation, 7290 coefficients would be needed for a complete representation. Using the rectified BWT, only 1458 coefficients are required. When using neurophysiology data to fit the STRF model, this saving in the number of coefficients substantially improves the quality of the resulting fits.

Compared to dyadic orthogonal transforms such as the Haar, triadic transforms such as the BWT have a further useful advantage for STRF estimation. Dyadic decomposition of space inevitably involves splitting images down the middle (Figure 6A, B): a centered STRF will be split into four sections, none of which is well-fit by one or a small number of wavelets. Thus, a dyadic decomposition fails to take advantage of the similarity between the structure of the STRF and the structure of the wavelets. In contrast, triadic decomposition separates the center of the STRF from its surroundings (Figure 6, resulting in a wavelet-like center section which can be well fit by a small number of BWT wavelets.

Several recent methods for STRF estimation (e.g. Theunissen et al. 2001; Sahani & Linden 2003; Willmore & Smyth 2003; David & Gallant 2005) have been developed for use with correlated stimuli, primarily natural scenes. These methods differ from classical techniques (Marmarelis & Marmarelis, 1978) in that they use regularization to produce STRF estimates that compensate for the statistical bias in complex stimuli. Regularization is a technique which balances *a priori* and *a posteriori* constraints to produce a STRF estimate which is both plausible and a good fit to experimental data. Successful regularized STRF estimation depends on choosing a basis set and an *a priori* constraint that are mutually compatible Wu et al. (2006). For example, in pixel space, V1 receptive fields tend to be smooth, and so a smoothness constraint may be appropriate. In terms of the BWT, however, V1 receptive fields tend to be sparse (i.e. they can be described using a small number of BWT coefficients). We have found the BWT and a sparseness constraint to be a useful combination for estimation of cortical STRFs.

5.2 Sparse image representation

Orthogonal (and other linearly independent) representations such as the BWT provide efficient bases for representing images because they minimize the total number of coefficients required to represent the image. However, this is not the only useful definition of 'efficiency'. Studies of the statistics of natural images have highlighted sparse coding as an alternative definition of efficiency (Field, 1994). Sparse codes minimize the number of *non-zero* coefficients required to represent information, rather than the *total* number.

The mammalian visual system may use sparse codes to represent the information present in natural retinal visual input (Rolls & Tovee, 1995; Baddeley et al., 1997; Vinje & Gallant, 2000). Neurons in V1 usually have high *lifetime* sparseness (sparse response of individual neurons to many inputs), since they represent sparsely-occurring wavelet structure in natural scenes (Olshausen & Field, 1996). The V1 population may also have high *population* sparseness (sparse response of the whole population to each input individually, Willmore & Tolhurst 2001); this would suggest that it forms a code where the redundancy between

neurons is minimized. It has been suggested that both lifetime sparseness (Olshausen & Field, 1996) and population sparseness (Field, 1994) are desirable for efficient coding.

Population and lifetime sparseness of the BWT representation were compared to several other codes for visual information, using methods described in Willmore & Tolhurst (2001). Several sets of basis functions were generated: (a) single pixels, (b) Difference-of-Gaussians, (c) Walsh patterns, (d) the Principal Components of natural scenes, (e) a Gabor pyramid, (f) a separable Haar pyramid, (g) a 2D Haar pyramid, and (h) a BWT pyramid. These were generated at 27×27 pixel resolution, except for the Walsh patterns and Haar pyramids, which were generated at 16×16 and 32×32 (since they are necessarily powers of 2 in size). The Gabor filters were similar to the pyramid proposed by Navarro et al. (1996): they were self-similar, even-symmetric filters with peak frequencies f_N, $f_N/2$, $f_N/4$ and $f_N/8$ (where f_N is the Nyquist frequency), oriented at $0°$, $45°$, $90°$, and $135°$, with spatial frequency bandwidth (ratio of half-maximum values) of 1 octave and full angular bandwidth at half-height of $41°$.

Basis functions that represented the DC were removed, and the remaining basis functions were standardized to zero mean and unit length. Sets of 2000 fragments of photographs of natural scenes were obtained at the appropriate sizes (after gamma correction), and the responses of each basis functions to each fragment were calculated. The fragments were then whited (by flattening of their power spectra), and the responses to these whitened fragments were calculated. Population sparseness of the responses of each basis set was measured by calculating the mean sparseness of the responses of the entire basis set to each image. Lifetime sparseness of each basis set was measured by calculating the mean sparseness of the responses of each basis function to the entire image set. In both cases, the following adaptation of the Treves-Rolls measure (Treves & Rolls, 1991; Vinje & Gallant, 2000) was used:

$$S = 1 - \frac{\left(\sum |r| \right)^2}{n \sum r^2} \quad (24)$$

Figure 7A shows the resulting sparseness values for the unwhitened images. It is well known that wavelet codes have high lifetime sparseness when used to represent natural images (Field, 1987), and this is confirmed by our observation that the Haar, BWT and Gabor bases have high lifetime sparseness compared to the other codes we investigated. High population sparseness results from several factors, including high lifetime sparseness and orthogonality (Willmore & Tolhurst, 2001). Here, we find that the orthogonal codes — the Haar bases, BWT, Principal Components and Walsh patterns — have high population sparseness. Figure 7B shows the sparseness values for whitened images. Whitening removes the difference in mean responsiveness between high- and low-frequency filters which results from the low-frequency bias in natural images. This reduces the population sparseness of most codes, especially the Principal Components and Walsh codes. The Haar and BWT codes remain the most lifetime-and population-sparse codes. Thus, the Haar and BWT are unusual in being orthogonal, lifetime-sparse and population-sparse. These codes minimize both the total number of coefficients required to represent images, and the number of these coefficients

that are non-zero. This combination of properties may be useful for practical image compression purposes.

6 Summary

The BWT is a wavelet basis for efficient representation of images. It shares many characteristics with the neural code for images in V1 and also has useful computational properties: it is complete, orthonormal, and a sparse code for natural images. As a model of neural coding in area V1, it is superior to most orthogonal transformations; in particular, it contains odd- and even-phase filters which are essential for modeling the phase invariant tuning of V1 complex cells. Although Gabor pyramids provide a better biological model, the orthogonality of the BWT is a significant advantage for some computational purposes.

Acknowledgments

References

Adelson EH, Bergen JR. Spatiotemporal energy models for the perception of motion. J Opt Soc Amer A. 1985; 2(2):284–299. [PubMed: 3973762]

Adelson EH, Simoncelli E, Hingorani R. Orthogonal pyramid transforms for image coding. Proc SPIE. 1987; 845:50–58.

Ahmed, OA., Fahmy, MM. Stable critically-sampled Gabor transform with localized biorthogonal function. Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis; 1998. p. 37-40.

Baddeley R, Abbott LF, Booth MCA, Sengpiel F, Freeman R, Wakeman EA, Rolls ET. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. Proc R Soc Lond B. 1997; 264:1775–1783.

Bell AJ, Sejnowski TJ. The "independent components" of natural scenes are edge filters. Vision Res. 1997; 37(23):3327–3338. [PubMed: 9425547]

Daubechies I. Orthonormal bases of compactly supported wavelets. Commun Pur Appl Math. 1988; 41:906–966.

Daugman JG. Two-dimensional spectral analysis of cortical receptive field profiles. Vision Res. 1980; 20(10):847–856. [PubMed: 7467139]

David SV, Gallant JL. Predicting neuronal responses during natural vision. Network: Comput Neural Syst. 2005; 16(2–3):239–260.

De Valois RL, Albrecht DG, Thorell LG. Spatial frequency selectivity of cells in macaque visual cortex. Vision Res. 1982; 22:545–559. [PubMed: 7112954]

Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Amer A. 1987; 4:2379–2394. [PubMed: 3430225]

Field DJ. What is the goal of sensory coding? Neural Comp. 1994; 6(4):559–601.

Gabor D. Theory of communication. J Inst Electr Eng. 1946; 93:429–457.

Haar A. Zur Theorie der orthogonalen Funktionensysteme. Mathematische Annalen. 1910; 69:331–371.

Jones JP, Palmer LA. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol. 1987; 58(6):1233–1258. [PubMed: 3437332]

Marcelja S. Mathematical description of the responses of simple cortical cells. J Opt Soc Amer. 1980; 70(11):1297–1300. [PubMed: 7463179]

Marmarelis, PZ., Marmarelis, VZ. Analysis of physiological systems: The white noise approach. Plenum; New York, NY: 1978.

Movshon JA, Thompson ID, Tolhurst DJ. Receptive field organization of complex cells in the cat's striate cortex. J Physiol. 1978a; 283:79–99. [PubMed: 722592]

Movshon JA, Thompson ID, Tolhurst DJ. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. J Physiol. 1978b; 283:53–77. [PubMed: 722589]

Navarro, R., Tabernero, A., Cristobal, G. Image representation with Gabor wavelets and its applications. In: Hawkes, PW., editor. Advances in Imaging and Electron Physics. Vol. 97. Academic Press; New York: 1996. p. 1-84.

Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996; 381(6583):607–609. [PubMed: 8637596]

Pollen DA, Ronner SF. Phase relationship between adjacent simple cells in the visual cortex. Science. 1981; 212(4501):1409–1411. [PubMed: 7233231]

Pollen DA, Ronner SF. Visual cortical neurons as localized spatial frequency filters. IEEE Trans Systems, Man and Cybernetics. 1983; 13:907–916.

Porat M, Zeevi YY. The generalized Gabor scheme of image representation in biological and machine vision. IEEE Trans Pattern Analysis and Machine Intelligence. 1988; 10(4):452–467.

Rao RM, Bundonis JS, Szu HH. Three-scale wavelet transforms. Proc SPIE. 1998; 3391:326–334.

Ringach DL, Shapley RM, Hawken MJ. Orientation selectivity in macaque V1: diversity and laminar dependence. J Neurosci. 2002; 22(13):5639–5651. [PubMed: 12097515]

Rolls ET, Tovee MJ. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol. 1995; 73(2):713–726. [PubMed: 7760130]

Sahani, M., Linden, JF. Evidence optimization techniques for estimating stimulus-response functions. In: Becker, S.Thrun, S., Obermayer, K., editors. Advances in Neural Information Processing Systems. Vol. 15. MIT Press; Cambridge, MA: 2003. p. 301-308.

Smyth D, Willmore B, Baker GE, Thompson ID, Tolhurst DJ. The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. J Neurosci. 2003; 23(11):4746–4759. [PubMed: 12805314]

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network: Comput Neural Syst. 2001; 12(3):289–316.

Treves A, Rolls E. What determines the capacity of autoassociative memories in the brain? Network: Comput Neural Syst. 1991; 2(4):371–397.

Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. Science. 2000; 287(5456):1273–1276. [PubMed: 10678835]

Watson AB. The cortex transform: rapid computation of simulated neural images. Comp Vis Graphics Image Proc. 1987; 39(3):311–327.

Watson, AB., Ahumada, AJ. A look at motion in the frequency domain. In: Tsotsos, JK., editor. Motion: Perception and representation. Association for Computing Machinery; New York: 1983. p. 1-10.

Watson AB, Ahumada AJ. A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex. IEEE Trans Biomed Eng. 1989; 36:97–106. [PubMed: 2921066]

Willmore B, Prenger RJ, Gallant JL. Spatial and temporal receptive field properties of neurons in area V2. Soc Neurosci Abstr. 2005; 618(17)

Willmore B, Prenger RJ, Gallant JL. Spatial and temporal receptive field properties of neurons in area V4. Soc Neurosci Abstr. 2006; 640(4)

Willmore B, Smyth D. Methods for first-order kernel estimation: simple-cell receptive fields from responses to natural scenes. Network: Comput Neural Syst. 2003; 14(3):553–577.

Willmore B, Tolhurst DJ. Characterizing the sparseness of neural codes. Network: Comput Neural Syst. 2001; 12(3):255–270.

Wu MC, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. Annu Rev Neurosci. 2006; 29:477–505. [PubMed: 16776594]

Appendix A Admissibility of the BWT

The admissibility condition is that:

$$C_\beta = \int_{-\infty}^{+\infty} \frac{|[\beta_{\theta,\phi}](\mathbf{k})|^2}{|\mathbf{k}|} d\mathbf{k} < \infty \tag{25}$$

where $\mathbb{F}[\cdot]$ is the Fourier transform. The BWT can be written as a linear combination of 2D square pulses (see Equation 4):

$$\beta_{\theta,\phi}(x,y) = \sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n} w(x-m, y-n) \tag{26}$$

Since the Fourier transform is linear, we can write:

$$[\beta_{\theta,\phi}(x,y)] = \sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n} [w(x-m, y-n)] \tag{27}$$

Using the triangle inequality, we can put an upper bound on the Fourier power spectrum of the BWT:

$$|[\beta_{\theta,\phi}(x,y)]|^2 = \left|\sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n} [w(x-m, y-n)]\right|^2 \tag{28}$$

$$\leq \sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n}^2 |[w(x-m, y-n)]|^2 \tag{29}$$

The Fourier transform of a translated function simply introduces a phase factor, with unit norm. Hence, the power spectrum is invariant to translation:

$$|[\beta_{\theta,\phi}(x,y)]|^2 \leq \sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n}^2 |[w(x,y)]|^2 \tag{30}$$

Using Fubini's theorem, the Fourier transform of the 2D pulse, $w(x, y)$, can be computed by the product of the respective Fourier transforms of 1D square pulses. The 1D square pulses, $u(x)$, can be written:

$$u(x) = \begin{cases} 1 & \text{if } 0 < x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

The Fourier transform of a 1D square pulse, $u(x)$, is well known and is given by the sinc function:

$$[u(x)](k) = \frac{1}{\sqrt{2\pi}} e^{-ik/2} \text{sinc}\left(\frac{k}{2}\right) \tag{32}$$

By Fubini's theorem, the Fourier transform of the 2D pulse, $w(x, y)$, is equal to a product of sinc functions:

$$\begin{aligned} [w(x,y)](\mathbf{k}) &= [u(x)](k_x) \cdot [u(y)](k_y) \\ &= e^{-ik_x/2} e^{-ik_y/2} \frac{2\sin(k_x/2)\sin(k_y/2)}{\pi k_x k_y} \end{aligned} \tag{33}$$

Thus, the Fourier power spectrum of $w(x, y)$ is:

$$|[w(x,y)](\mathbf{k})|^2 = \frac{4\sin^2(k_x/2)\sin^2(k_y/2)}{\pi^2 k_x^2 k_y^2} \tag{34}$$

Using Equations 25 and 29, we can put an upper bound on the admissibility constant:

$$C_\beta \equiv \int_{-\infty}^{+\infty} |[\beta_{\theta,\phi}](\mathbf{k})|^2 |\mathbf{k}| d\mathbf{k} \tag{35}$$

$$\leq \int_{-\infty}^{+\infty} \frac{\sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n}^2 |[w(x,y)]|^2}{|\mathbf{k}|} d\mathbf{k} \tag{36}$$

$$= \sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n}^2 \int_{-\infty}^{+\infty} \frac{|[w(x,y)]|^2}{|\mathbf{k}|} d\mathbf{k} \tag{37}$$

$$= \sum_{m=1}^{3}\sum_{n=1}^{3} a_{m,n}^2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{4\sin^2(k_x/2)\sin^2(k_y/2)}{\pi^2 k_x^2 k_y^2 \sqrt{k_x^2+k_y^2}} dk_x dk_y \tag{38}$$

The double integral in equation 38 is finite, and given by:

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} \frac{4\sin^2(k_x/2)\sin^2(k_y/2)}{\pi^2 k_x^2 k_y^2 \sqrt{k_x^2+k_y^2}} dk_x dk_y = \frac{1}{24\pi}\left[52-40\sqrt{2}+8\sinh^{-1}(1)\right]+\frac{32}{24\pi}\left[\sinh^{-1}\left(\sqrt{1/\sqrt{2}-1/2}\right)+{}_3F_2\left(1,1,\tfrac{3}{2};2,3;-1\right)+12\ln(4)\right.$$

$$\simeq 0.4732$$

(39)

where ${}_pF_q$ is the generalized hypergeometric function. Thus, the admissibility constant, C_β, for the BWT wavelets is finite, and the BWT wavelets are admissible wavelets.

Appendix B Orthogonality of the BWT

We can show that the BWT is orthogonal by making use of the symmetry of the basis functions. We define three coordinate transforms for a function, $f(x, y)$:

$$\mathbb{I}_x(f):f(x,y) \to f(-x,y) \quad (40)$$

$$\mathbb{I}_y(f):f(x,y) \to f(x,-y) \quad (41)$$

$$\mathbb{R}(f):f(x,y) \to f(-y,x) \quad (42)$$

$\mathbb{I}_x$ and $\mathbb{I}_y$ are coordinate inversion operators that invert the x and y coordinates respectively. Applying either operator twice leaves both coordinates unchanged, i.e. $\mathbb{I}_x^2 = \mathbb{I}_y^2 = \mathbb{I}$, where $\mathbb{I}$ is the identity operator. $\mathbb{R}$ is an operator that rotates the coordinates clockwise by 90°. Applying $\mathbb{R}$ twice gives a 180° rotation. From the definitions of the operators, one can derive the identity $\mathbb{R}^2 = \mathbb{I}_x\mathbb{I}_y = \mathbb{I}_y\mathbb{I}_x$.

The even-phase wavelets, $\beta_{\theta,e}$ are symmetric under $\mathbb{R}^2$. The odd-phase wavelets, $\beta_{\theta,o}$, are anti-symmetric under $\mathbb{R}^2$. Also, the vertically oriented wavelets, $\beta_{|,\phi}$, are symmetric under $\mathbb{I}_y$ and the horizontally oriented wavelets, $\beta_{-,\phi}$, are symmetric under $\mathbb{I}_x$.

The inner product of two functions can be defined as:

$$\langle f,g\rangle = \int\limits_{X\times Y} f(x,y)g(x,y)\,dxdy \quad (43)$$

Since integration is invariant to any coordinate transform, the inner product operator, $\langle\cdot,\cdot\rangle$, is invariant to the transforms defined in equations 40, 41 and 42. Thus, we can show that any

even-phase wavelet is orthogonal to any odd-phase wavelet by writing the inner product between them as:

$$\langle \beta_{\theta,e}, \beta_{\theta,o} \rangle = \langle \mathbb{R}^2 \beta_{\theta,e}, \mathbb{R}^2 \beta_{\theta,o} \rangle = \langle \beta_{\theta,e}, -\beta_{\theta,o} \rangle = -\langle \beta_{\theta,e}, \beta_{\theta,o} \rangle \quad (44)$$

Since the only real number that is equal to its negative is zero, the inner product between any even-phase wavelet and any odd-phase wavelet is zero.

We can show that the vertical wavelets are orthogonal to the horizontal wavelets by separating the integral that defines their inner product into two separate integrals in *x* and *y*. Since each wavelet has zero mean, the inner product must be zero:

$$\begin{aligned}\left\langle b_{|,\phi}, b_{-,\phi} \right\rangle &= \underset{X \times Y}{\int} b_{|,\phi}(x,y) b_{-,\phi}(x,y)\, dx dy \\ &= k \underset{X}{\int} b_{|,\phi}(x)\, dx \underset{Y}{\int} b_{-,\phi}(y)\, dy \\ &= 0 \end{aligned} \quad (45)$$

Similarly, for the other orientations, we can show:

$$\left\langle b_{|,\phi}, b_{/,\phi} \right\rangle = \left\langle \mathbb{I}_y b_{|,\phi}, \mathbb{I}_y b_{/,\phi} \right\rangle = \left\langle b_{|,\phi}, b_{\backslash,\phi} \right\rangle \quad (46)$$

$$\left\langle b_{|,\phi}, b_{/,\phi} \right\rangle = \left\langle \mathbb{R} b_{|,\phi}, \mathbb{R} b_{/,\phi} \right\rangle = \left\langle b_{|,\frac{e}{o}}, \pm b_{/,\frac{e}{o}} \right\rangle = \pm \left\langle \mathbb{I}_x b_{-,\frac{e}{o}}, \mathbb{I}_x b_{\frac{e}{o}} \right\rangle = \pm \left\langle b_{-,\frac{e}{o}}, \pm b_{/,\frac{e}{o}} \right\rangle \quad (47)$$

$$\langle b_{-,\phi}, b_{/,\phi} \rangle = \left\langle \mathbb{I}_y b_{|,\phi}, \mathbb{I}_y b_{/,\phi} \right\rangle = \left\langle b_{|,\phi}, b_{\backslash,\phi} \right\rangle \quad (48)$$

Appendix C Uniqueness of the BWT

The BWT mother wavelets are uniquely specified by the following criteria:

- Defined on a 3×3 grid

- Orthonormal basis (i.e. orthogonal, complete, unit length)

- There is a single basis function representing D.C.

- Every other wavelet varies along only one axis (0°, 45°, 90° or 135°)

- There is an odd-even pair at each orientation

To demonstrate this, we can first specify the wavelets algebraically, taking into account the orientation and symmetry constraints:

$$b_0 = \begin{bmatrix} a & a & a \\ a & a & a \\ a & a & a \end{bmatrix} \quad (49)$$

$$b_{|,o} = \begin{bmatrix} -b & 0 & b \\ -b & 0 & b \\ -b & 0 & b \end{bmatrix} \quad (50)$$

$$b_{|,e} = \begin{bmatrix} c & d & c \\ c & d & c \\ c & d & c \end{bmatrix} \quad (51)$$

$$b_{/,o} = \begin{bmatrix} e & f & 0 \\ f & 0 & -f \\ 0 & -f & -e \end{bmatrix} \quad (52)$$

$$b_{/,e} = \begin{bmatrix} g & h & i \\ h & i & h \\ i & h & g \end{bmatrix} \quad (53)$$

$$b_{-,o} = \begin{bmatrix} -j & -j & -j \\ 0 & 0 & 0 \\ j & j & j \end{bmatrix} \quad (54)$$

$$b_{-,e} = \begin{bmatrix} k & k & k \\ l & l & l \\ k & k & k \end{bmatrix} \quad (55)$$

$$b_{\backslash,o} = \begin{bmatrix} 0 & m & n \\ -m & 0 & m \\ -n & -m & 0 \end{bmatrix} \quad (56)$$

$$b_{\backslash,e} = \begin{bmatrix} p & q & r \\ q & p & q \\ r & q & p \end{bmatrix} \qquad (57)$$

By applying the various constraints above, we can obtain a single, unique value for each parameter a, b, c … q, ignoring sign flips which merely imply a corresponding sign flip of single basis functions. These values correspond to the BWT.

The unit length constraint for b_0 specifies that $9a^2 = 1$, hence $a = 1/\sqrt{9}$. Note that to maintain orthogonality with b_0, all other basis functions must therefore have zero D.C.

For $b_{|,o}$, applying the unit length constraint determines that $6b^2 = 1$, and so $b = 1/\sqrt{6}$. We will select the positive or negative square roots arbitrarily, since they merely imply a sign flip of each basis function. Similarly for $b_{-,o}$, $j = 1/\sqrt{6}$.

The structure of $b_{|,e}$ is determined by the zero D.C. and unit length constraints: $6c + 3d = 0$ and $6c^2 + 3d^2 = 1$, hence $d = -2c$ and so $c = -1/\sqrt{18}, d = 2/\sqrt{18}$. Similarly for $b_{-,e}$, $k = -1/\sqrt{18}$ and $l = 2/\sqrt{18}$.

To determine the structure of the odd oblique wavelets, we need to invoke the orthogonality constraint. For example, $b_{/,o}$ must be orthogonal to $b_{|,o}$, and so $-eb - fb - gb + gb - fb - eb = 0$, $\therefore e = -f$. Then, using the length constraint, $6e^2 = 1$ and so $e = -1/\sqrt{6}$ and $f = 1/\sqrt{6}$. Similarly for $b_{\backslash,o}$, $m = -1/\sqrt{6}$ and $n = 1/\sqrt{6}$.

The even oblique wavelets are also specified by the orthogonality constraint. For example, $b_{/,e}$ must be orthogonal to $b_{|,e}$, and so $c(g + h + i) + d(2h + i) + c(g + h + i) = 0$. Since $d = -2c$, $2c(g + h + i) - 2c(2h + i) = 0$, and so $g = h$. Now, applying the zero D.C. and unit length constraints, $3g + 3h + 3i = 0$ and $3g^2 + 3h^2 + 3i^2 = 1$. Substituting for h, we obtain $6g + 3i = 0$ and $6g^2 + 3i^2 = 1$, and so $g = h = -1/\sqrt{18}$ and $i = 2/\sqrt{18}$. Similarly for $b_{\backslash,e}$, $p = 2/\sqrt{18}$ and $q = r = -1/\sqrt{18}$.
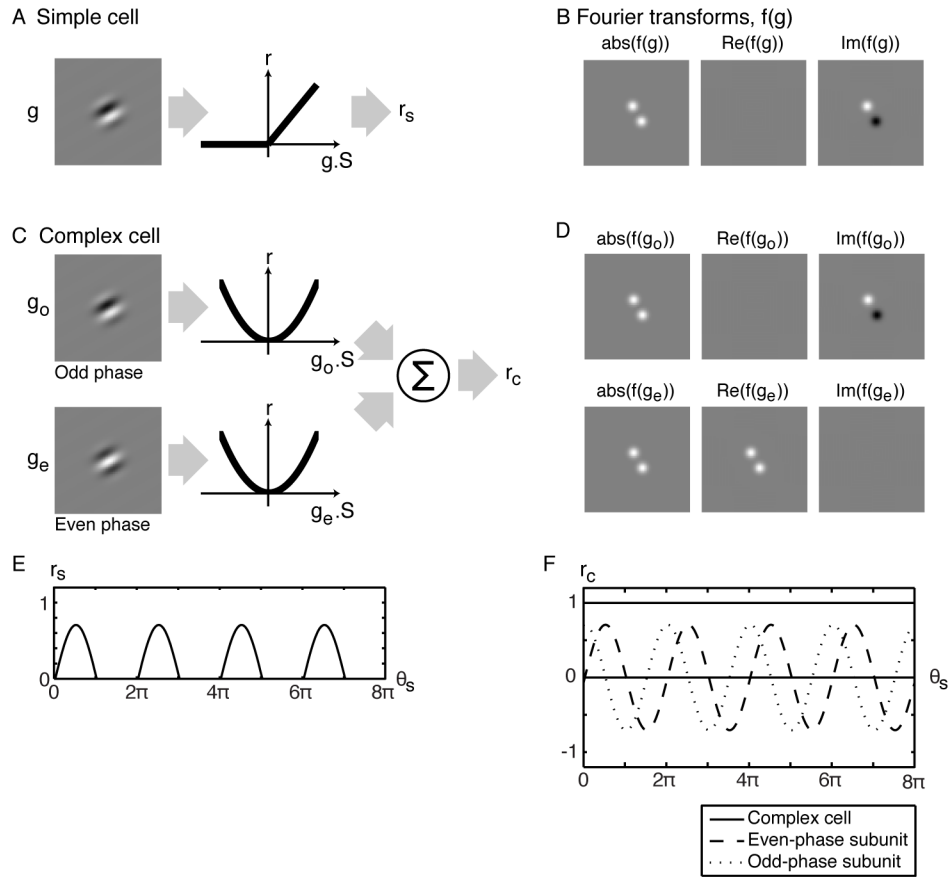
Figure 1.
Standard models of cortical simple and complex cells. A. Simple cells can be modeled using an LN model, consisting of a 2D Gabor wavelet SRF, followed by half-wave rectification. In this case, a Gabor with odd phase is shown; real simple cells have a variety of phase preferences. B. Fourier transform of the SRF in A. The SRF is band-pass in spatial-frequency and orientation, and odd-symmetric. C. Complex cells can be modeled as the sum of the squares of the responses of a pair of 2D Gabor wavelets in quadrature phase. D. Fourier transform of the wavelets in C, showing that they are a quadrature pair with odd and even phase. E. Response of the model simple cell to sinusoidal stimuli with the same spatial-frequency and orientation tuning as the SRF, but varying in phase. The response is strongly modulated by phase variation. F. Response of the model complex cell to phase variation. The two Gabor components (dotted and dashed lines) are modulated by phase, but the sum of the squares of their responses (solid line) is phase invariant.
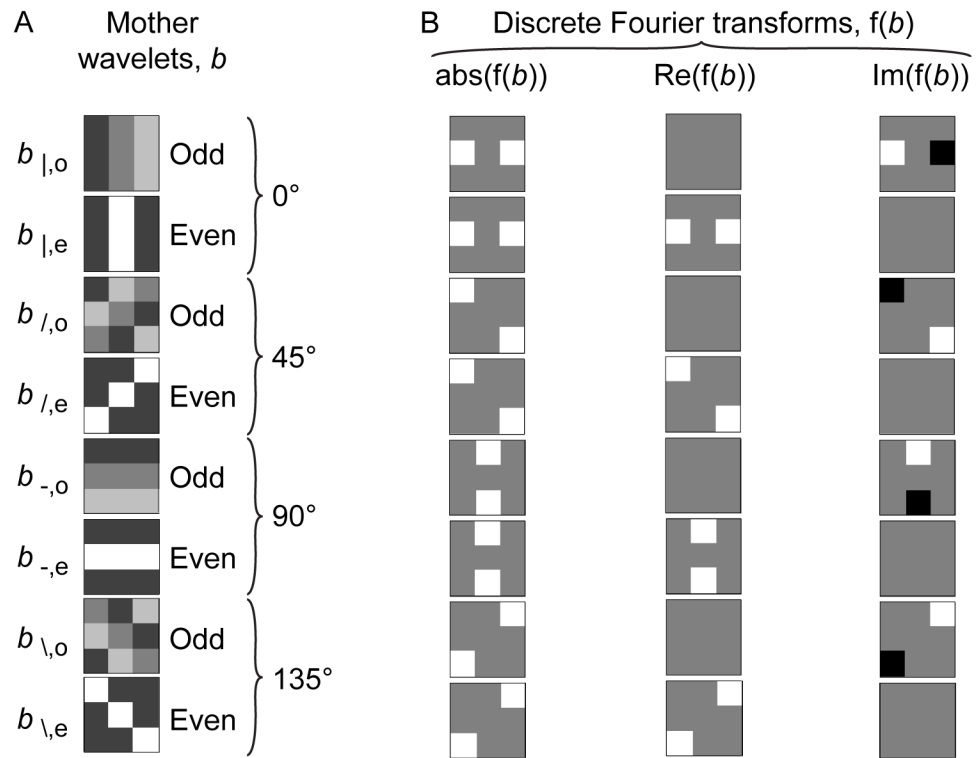
Figure 2.

The eight BWT mother wavelets, showing how the transform was constructed. A. The eight wavelets in pixel space. B. 3 × 3 Discrete Fourier Transforms (DFTs) of the mother wavelets. The wavelets were chosen so that each DFT occupies a single pair of pixels (as shown by the absolute values), because this minimises their spatial-frequency and orientation bandwidths as far as is possible within this space. If the set is also to be orthogonal, this constraint requires that there will be four pairs of wavelets, at four different orientations (0°, 45°, 90°, 135°). The pairs were chosen so that each contains one wavelet with an even-symmetric real DFT, corresponding to an even-symmetric function in pixel space, and one wavelet with an odd-symmetric imaginary DFT, corresponding to an odd-symmetric function in pixel space. The 3 × 3 DFTs are qualitatively similar to those of the Gabor filters in Figure 1. When analysed at higher resolution, the wavelets are less well localised in Fourier space (Figure 4).
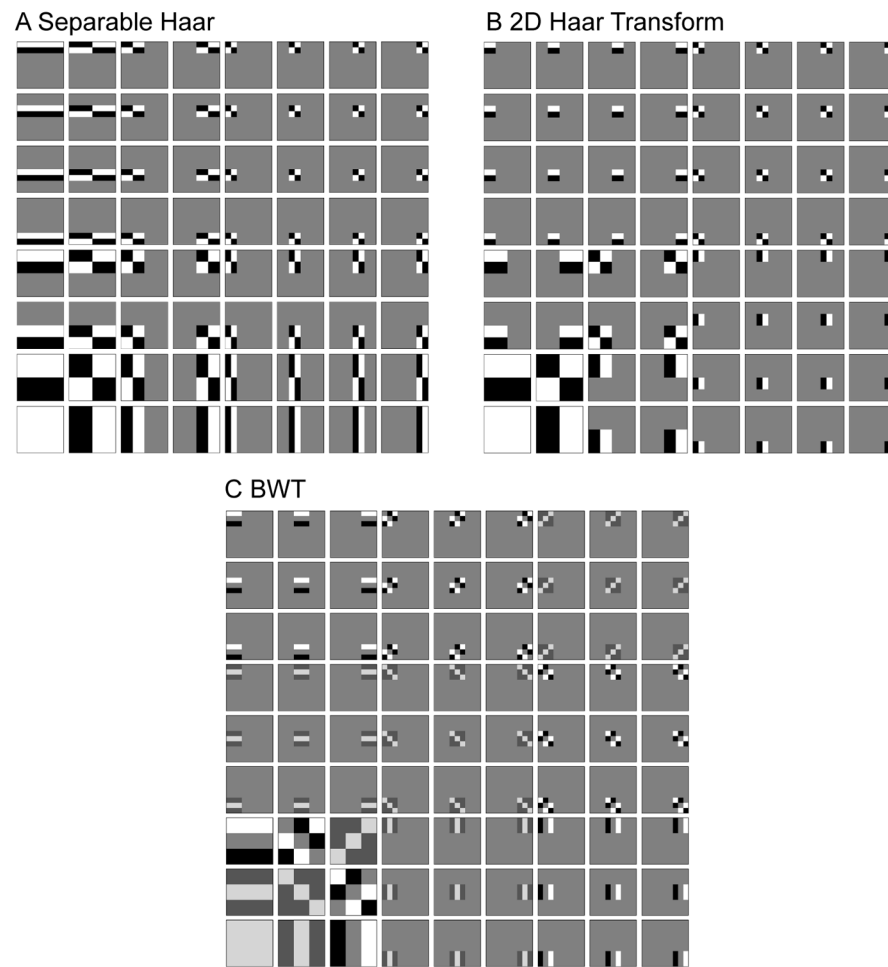
A Separable Haar

B 2D Haar Transform

C BWT

Figure 3.
Comparison of the BWT with 2D generalizations of the Haar transform. A. Separable Haar transform produced by separable combination of all pairs of 1D Haar wavelets in x and y. The resulting wavelets are asymmetrically stretched versions of the mother wavelets, and so the transform is not self-similar. B. 2D Haar transform produced by separable combination of the mother wavelets, followed by translation and dilation. The resulting set is approximately self-similar, and approximately Gabor-like; however all wavelets have odd phase. C. BWT. The set is similar across scale, and roughly similar across orientation, and is composed of odd-even pairs.
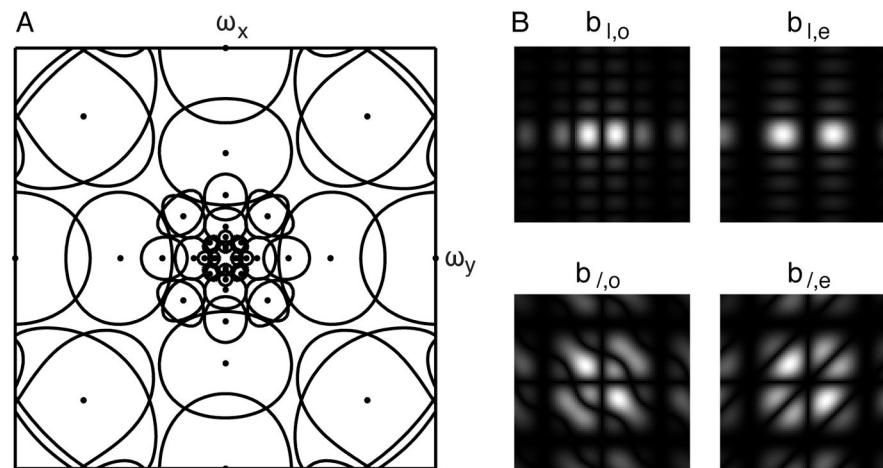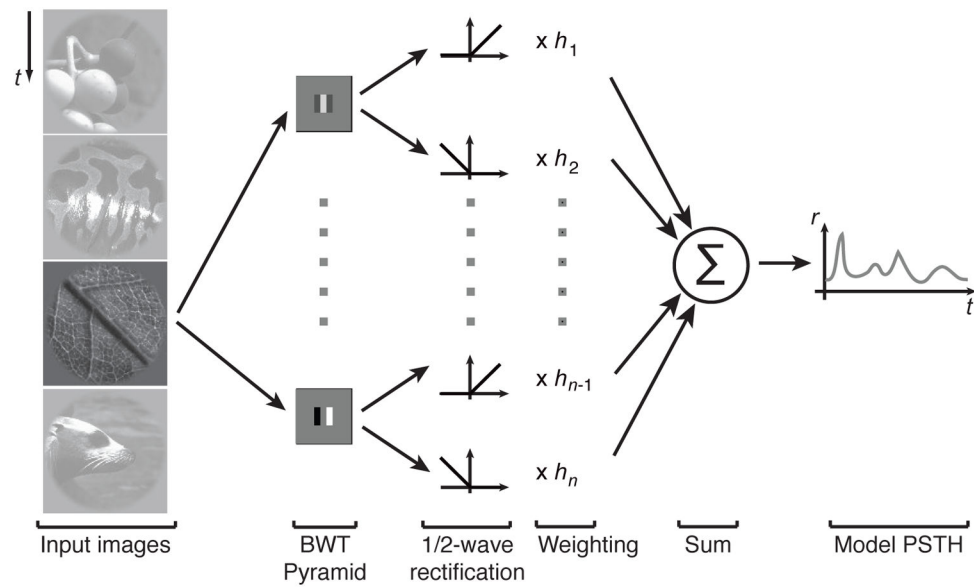
Figure 4.
Frequency tuning characteristics of the BWT basis. A. Contour plot showing the Fourier power spectra of three complete frequency bands of the BWT. Each dot indicates the peak of the tuning curve of one wavelet. Each surrounding contour marks the line where the tuning curve of that wavelet has dropped to $1/\sqrt{2}$ of its peak. The wavelets tile frequency space neatly, although the spatial frequency preferences of the wavelets within each frequency band are not identical, and the tuning curves are not identical in shape. At this criterion, each wavelet has a single closed contour, showing that the wavelets are fairly well localised in Fourier space. B. Power spectra of four BWT wavelets, showing some problems with spectral localisation. The sharp edges of the axially-aligned wavelets produce some tuning for harmonics of the central frequency, and the oblique wavelets show some tuning for cross-orientation structure. The other four wavelets are rotations of those shown.

Figure 5.
STRF model incorporating the BWT. Each image is passed through a complete BWT transform (of which two filters are shown), and the resulting responses are half-wave rectified, taking the positive and negative responses separately. To describe the responses of a single neuron, the responses are weighted and summed to produce a model PSTH. The weights, h_i, constitute a nonlinear STRF in wavelet space.
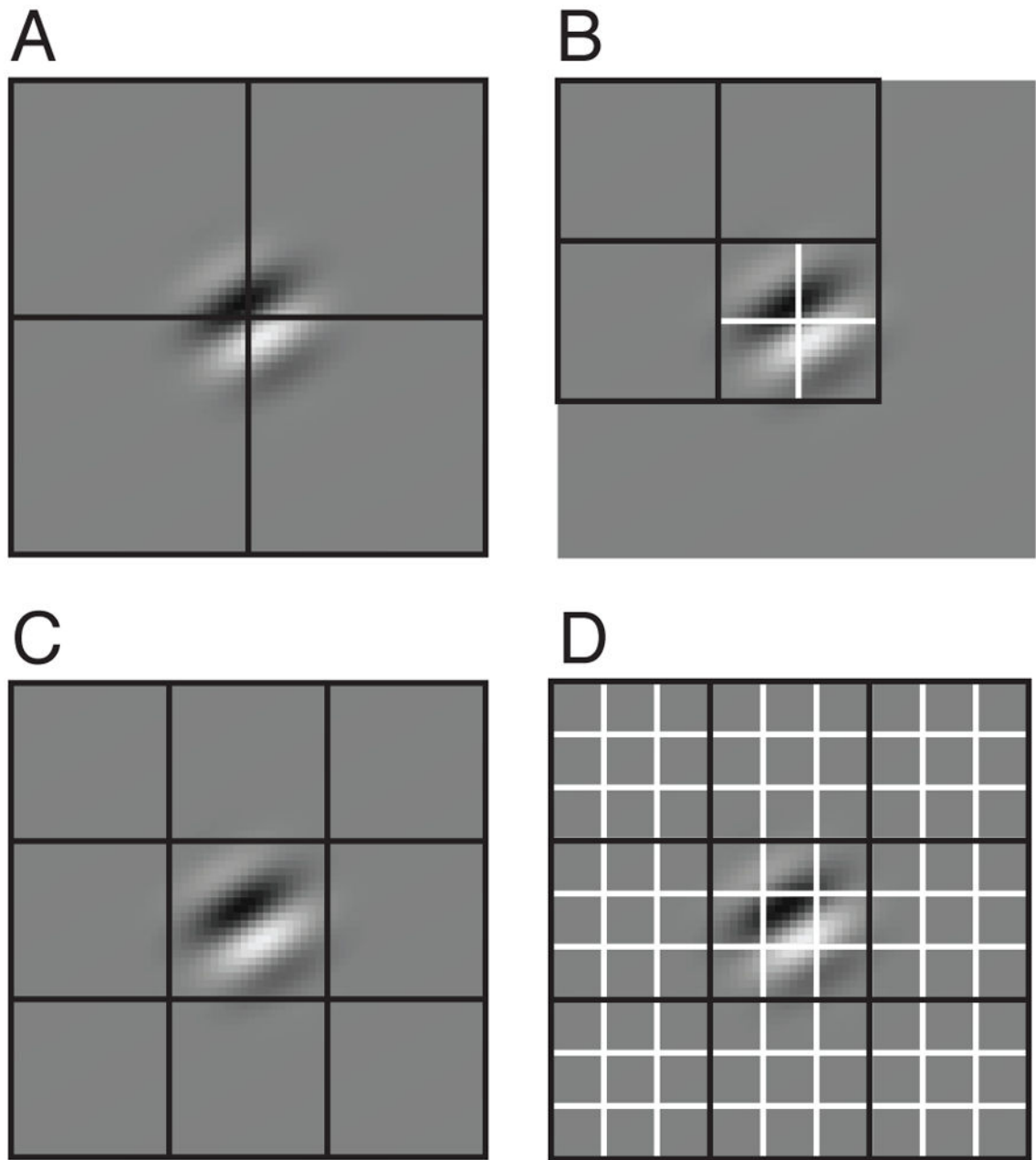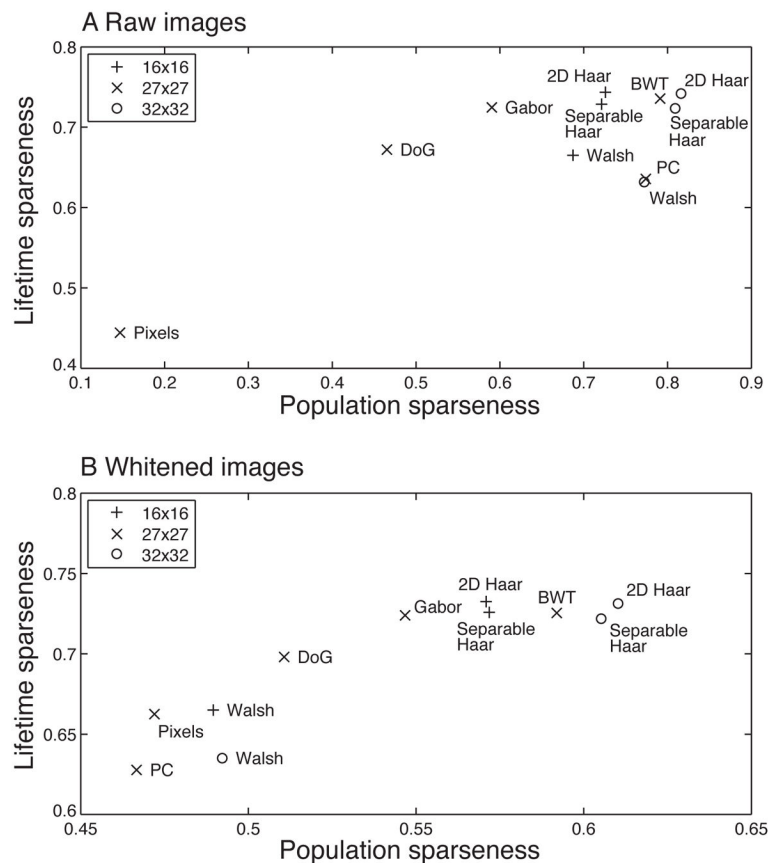
Figure 6.
Comparison of spatial tiling of dyadic and triadic wavelet decompositions. A. Consider a neural STRF, represented here by a Gabor filter. Dyadic division of space cuts the STRF down the center. This splits it into four sections, none of which is well described by a single wavelet. B. Even if the first division is offset (black grid) so that the STRF is not cut, subsequent division (white grid) still have the same effect. C. Triadic division of the same STRF separates the STRF from its surroundings. The central square contains most of the structure of the STRF, and can be well fit using a small number of wavelets. D. Further triadic divisions again focus on the center of the STRF.

Figure 7.
Sparse coding of natural images using the BWT. A. We used the BWT, and several common basis sets to encode a set of linearized natural images, and compared the population sparseness and lifetime sparseness of the responses of each basis. Orthogonal codes containing basis functions at a range of spatial frequencies have high population sparseness (BWT, Two-dimensional Haar, Separable Haar, Principal Components, Walsh patterns). Wavelet codes have high lifetime sparseness (BWT, Two-dimensional Haar, Separable Haar, Gabor). The BWT and Haar codes have high population and lifetime sparseness. B. To ensure that the high population sparseness of the wavelet codes does not result simply from the low-frequency bias of natural scenes, we repeated the analysis on images that had had their power spectra flattened. This reduces the population sparseness of most codes, especially the Principal Components and Walsh bases. The BWT and Haar codes still have the highest lifetime and population sparseness.