Natural Scenes, Vision and Wavelets

Nick Kingsbury

Signal Processing and Communications Laboratory
Department of Engineering, University of Cambridge, UK.

email: ngk@eng.cam.ac.uk web: www.eng.cam.ac.uk/~ngk

CSP Forum, Jan 2013

UNIVERSITY OF CAMBRIDGE

Natural Scenes, Vision and Wavelets

Some questions we shall try to answer:

- What are the characteristics of natural scenes?
- Why is human vision related to them?
- How is human vision related to them?
- What are wavelets?
- What are dual-tree complex wavelets?
- How do wavelets relate to the human visual system?
- Why are dual-tree wavelets good for computer vision systems?

Natural Scene 1 (Olympic Temperate Rain-forest, WA, USA)

Natural Scene 2 (Olympic Temperate Rain-forest, WA, USA)

Natural Scene 3 (Olympic Coast, WA, USA)

What are the characteristics of natural scenes?

- Lots of textures
- Large objects and small objects (many different scales)
- Lots of edges separating regions of different colours or intensities
- Not many straight edges
- Lots of fine detail that is not very important (except when it is food)
- Most things are stationary, but anything that moves could be dangerous!

Why is human vision related to natural scenes?

- For hundreds of thousands of years mammals and then humans **evolved in a natural landscape**.
- The mammalian vision system was a **key survival mechanism**, finding parents, food and mates, and warning of predators.
- We believe the **early stages of the vision system** – the eye, the retina and the V1 cortex at the back of the head – **all evolved so as to make subsequent processing of images as efficient as possible**.
- Key tasks were **recognition** of food, mates and other objects of importance for survival, such as predators or places to shelter.
- **Motion detection** was a key feature of predator detection and of recognition of potential sources of food.

How is human vision related to natural scenes?

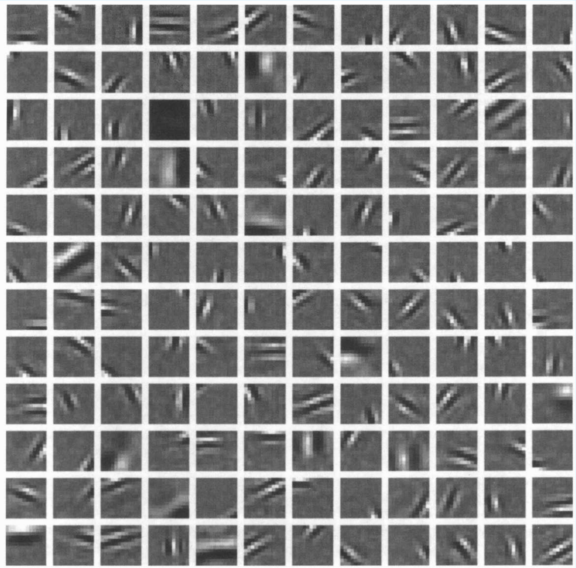
- The human eye, with its vari-focal lens and retina, is capable of taking in detail of scenes with a peak resolution corresponding to **over a million pixels** (picture elements) in the whole scene.
- The V1 cortex, at the far end of the optic nerve bundle from each eye (near the rear of the head), analyses the image from each retina and converts it into many millions of **sparsely coded neural pulses** which feed on to higher levels of brain function.
- Sparse coding is the **key to efficiency** at the inputs to these higher levels, because it only generates neural activity when there is something of potential importance at a given part of the image.
- Each group of neurons in the V1 cortex is connected to a local group of receptors on the retina in such a way that it **selects out** or **filters** a specific oriented pattern of a certain size or scale. Fine-scale patterns tend to be very small; coarse-scale patterns are larger.

How is human vision related to natural scenes? (cont.)

- The cortical filters perform a task that is equivalent to **correlating the image with many different patterns**, at various scales, orientations and locations; and the filter outputs from V1 are the results of all these correlations.
- **Bruno Olshausen** and **David Field** of Cornell showed that it is possible to **learn these sparsity-inducing patterns directly from natural scenes** in their 1996 Nature paper 'Emergence of simple-cell receptive field properties by learning a sparse code for natural images'.
- **They showed that the patterns, learnt from natural scenes by imposing a sparse coding requirement, were remarkably similar to the patterns that we find in the human V1 cortex that have evolved over hundreds of thousands of years of mammalian evolution.**

Sparsity-inducing Patterns learnt from Natural Scenes

Olshausen and Field's sparsity-inducing patterns, learnt directly from natural scenes, by finding the 144 patterns of 12×12 pixels, which can represent the scenes with as few non-zero coefs as possible.



Knowledge about the V1 cortex from Neurophysiology

- **David Hubel and Torsten Wiesel** won the Nobel prize in 1981 for their work, published around 1962, on characterising the V1 cortex of a cat. They anaesthetised a cat, inserted a fine electrode into the V1 cortex in various places, and oberved responses on the electrode to spots of light shone onto a screen in front of the cat. Hubel and Wiesel's results in fig 2 of their 1962 paper (page 6) show **positive (excitatory) and negative (inhibitory) responses, arranged in small 'stripy' patterns** at a full range of orientations and a range of scales.
- **Horace Barlow** has studied the mammalian and human visual systems from 1943 to the present day, and has published around 200 journal papers in this time, many of which are downloadable from the Trinity website: **www.trin.cam.ac.uk\horacebarlow** . Horace's interests have been very wide-ranging and have covered **the retina, the V1 cortex, higher level vision functions, motion detection and inference of general brain functionality**.
- Another interesting Trinity figure is **David Marr** who wrote a seminal book 'Vision' from the point of view of Computation and Physiology, just before dying tragically of leukemia in 1980.

Hubel and Wiesel, 1962 – Receptive Fields in the Cat's Visual Cortex

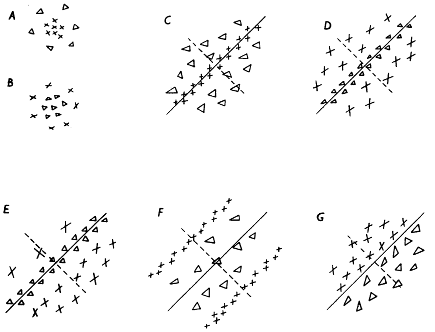
106 *J. Physiol.* (1962), **160**, *pp.* 106–154
With 2 plates and 20 text-figures
Printed in Great Britain

RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX

By D. H. HUBEL and T. N. WIESEL
From the Neurophysiolo.jy Laboratory, Department of Pharmacology Harvard Medical School, Boston, Massachusetts, U.S.A.

(*Received* 31 *July* 1961)

What chiefly distinguishes cerebral cortex from other parts of the central nervous system is the great diversity of its cell types and interconnexions. It would be astonishing if such a structure did not profoundly modify the response patterns of fibres coming into it. In the cat's visual cortex, the receptive field arrangements of single cells suggest that there is indeed a degree of complexity far exceeding anything yet seen at lower levels in the visual system.

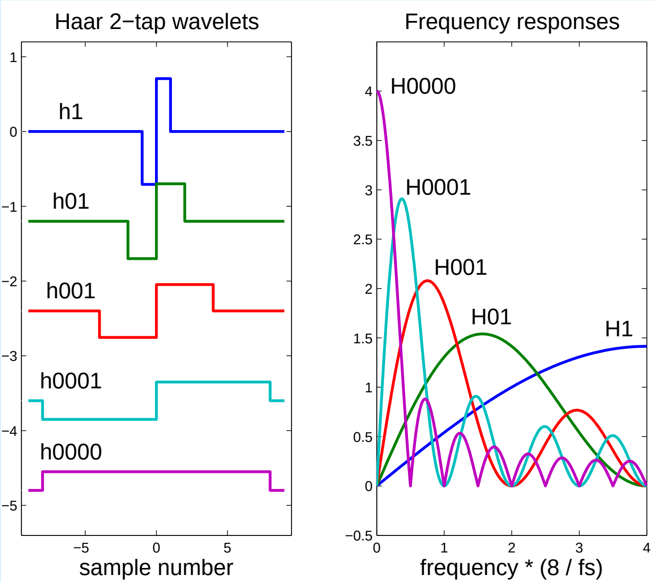


Text-fig. 2. Common arrangements of lateral geniculate and cortical receptive fields. *A*. 'On'-centre geniculate receptive field. *B*. 'Off'-centre geniculate receptive field. *C–G*. Various arrangements of simple cortical receptive fields. ×, areas giving excitatory responses ('on' responses); △, areas giving inhibitory responses ('off' responses). Receptive-field axes are shown by continuous lines through field centres; in the figure these are all oblique, but each arrangement occurs in all orientations.

Wavelets - what are they?

- Wavelets are the basis functions of a relatively new type of mathematical transform (since the early 1980s), that occupies a space somewhere between the spatial domain of image pixels, and the frequency (Fourier) domain of spatial frequency components.
- Instead of using pure cosine and sine waves, as Fourier does, wavelet functions are scaled and shifted versions of a common **mother wavelet** shape. In that sense they are simpler than Fourier functions. Usually wavelets are scaled in **generations**: each parent having 4 children in 2-D, and each child being half the size (in length) of its parent. Each child then has 4 grandchildren etc., for typically 4 to 8 generations.

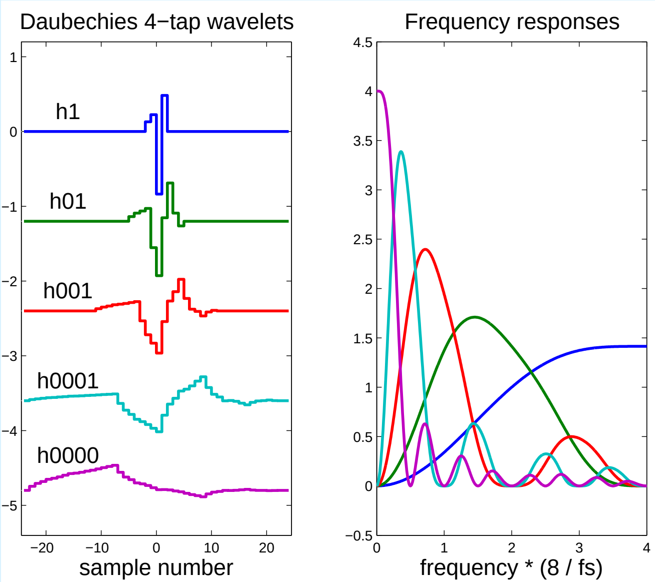
Very simple wavelets (2-tap Haar)



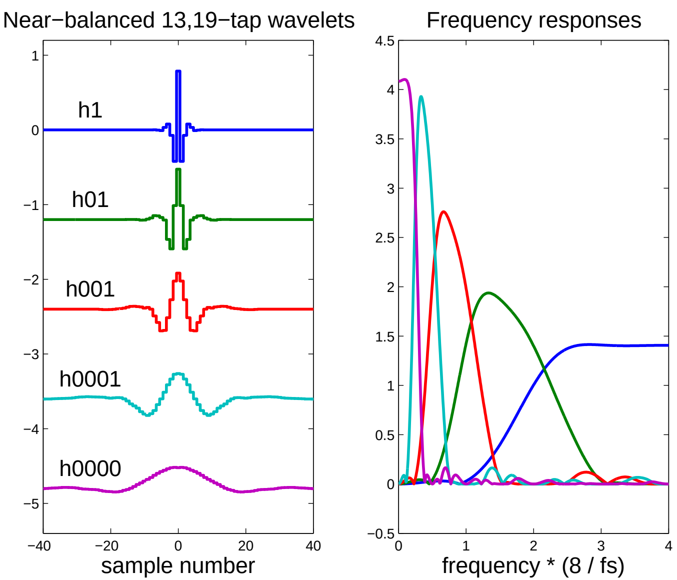
Wavelets - what are they? (cont.)

- The **inverse discrete wavelet transform (DWT)** builds an image up from the sum of a large number of the wavelet functions, each multiplied by an appropriate wavelet coefficient to generate the desired image.
- The **forward DWT** analyses an image with filters in order to calculate the correct values for all of the wavelet coefficients.
- The DWT is derived from a simple set of mathematical axioms, involving **self-similarity across scales** and **orthogonality**.

Simple non-symmetric wavelets (4-tap Daubechies)



Smooth symmetric wavelets (13,19-tap Tay-Kingsbury)



Real Discrete Wavelet Transform (DWT) in 1-D

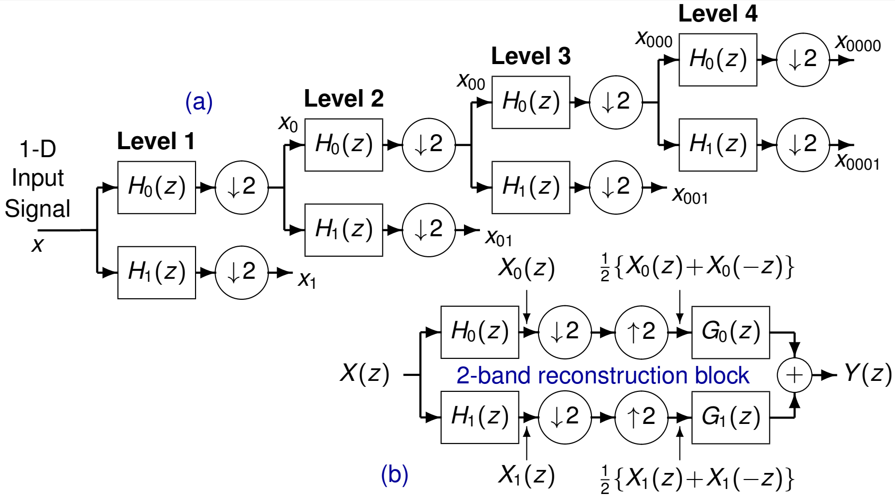


Figure: (a) Tree of real filters for the DWT. (b) Reconstruction filter block for 2 bands at a time, used in the inverse transform.

Discrete Wavelet Transform Features

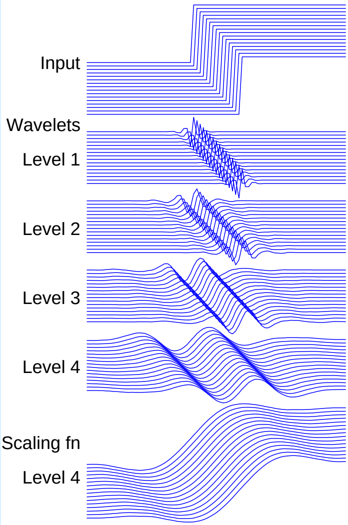
Features of the (Real) Discrete Wavelet Transform (DWT):

- **Good compression** of signal energy into sparse sets of coefficients.
- **Perfect reconstruction** with short support filters.
- **No redundancy**.
- **Very low computation** – order-N only.

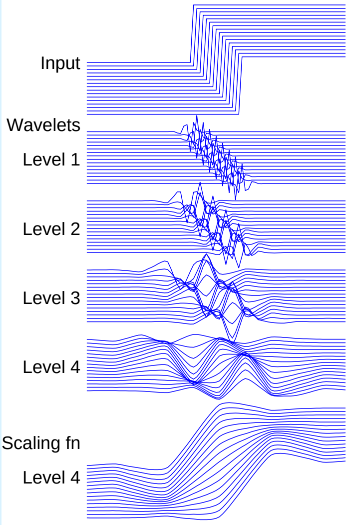
But what are the problems of the DWT?

- **Severe shift dependence** (due to aliasing in down-samplers).
- **Poor directional selectivity** in 2-D, 3-D etc. (due to separable real filters).

Shift Invariance of Complex DT-CWT vs Real DWT



(a) Dual Tree CWT



(b) Real DWT

What are dual-tree complex wavelets?

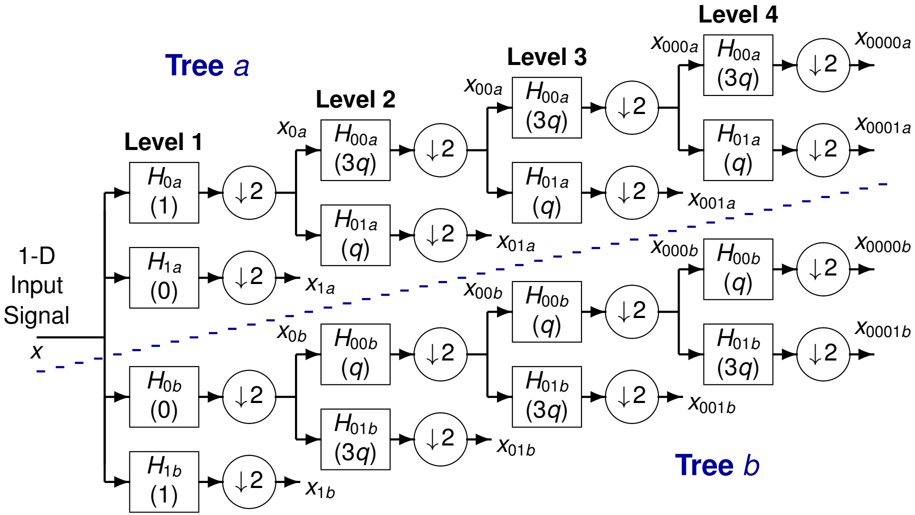
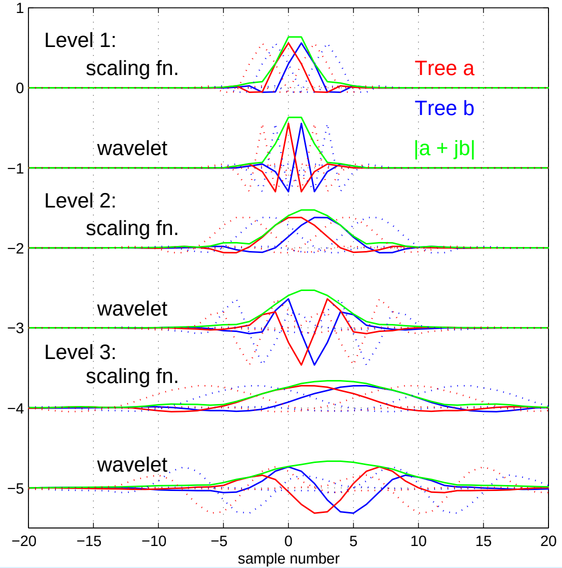


Figure: Dual tree of real filters for the Q-shift CWT, giving real and imaginary parts of complex coefficients from tree *a* and tree *b* respectively.

Q-shift DT CWT Basis Functions – Levels 1 to 3

Basis functions for adjacent sampling points are shown dotted.

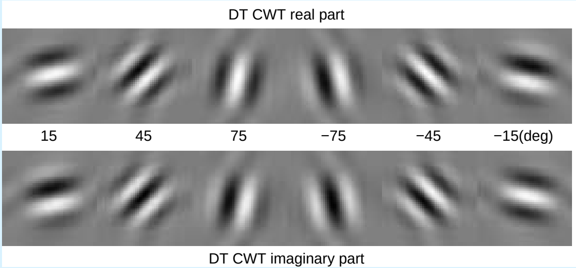


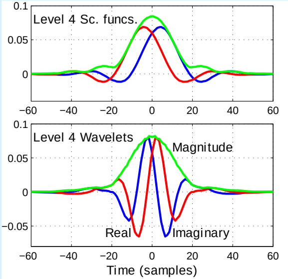
How do we extend the DT-CWT to multi-dimensions?

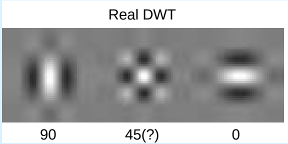
When the DT-CWT is applied to 2-D signals (images), it has the following features:

- It is performed separably, using 2 trees for the rows of the image and 2 trees for the columns – yielding a **Quad-Tree** structure (4:1 redundancy).
- The 4 quad-tree components of each coefficient are combined by simple sum and difference operations to yield a **pair of complex coefficients**. These are part of two separate subbands in adjacent quadrants of the 2-D spectrum.
- This produces **6 directionally selective subbands** at each level of the 2-D DT CWT. Fig 3 shows the basis functions of these subbands at level 4, and compares them with the 3 subbands of a 2-D DWT.
- The DT-CWT is directionally selective because the complex filters can **separate positive and negative frequency components** in 1-D, and hence **separate adjacent quadrants** of the 2-D spectrum. Real separable filters cannot do this!

Why do we get good directional filters in 2-D?





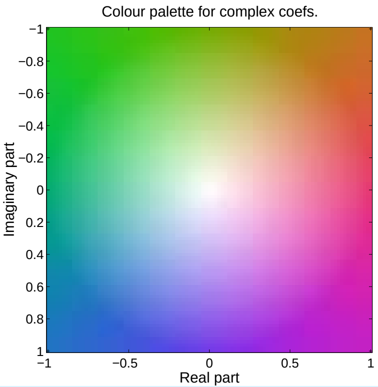


$$g(x)\, e^{i\omega_1 x} \,.\, g(y)\, e^{i\omega_2 y} =$$

$$g(x)\, g(y)\; e^{i(\omega_1 x + \omega_2 y)}$$

Figure: Basis functions of 2-D Q-shift complex wavelets (top), and of 2-D real wavelet filters (bottom), all illustrated at level 4 of the transforms. The complex wavelets provide 6 directionally selective filters, while real wavelets provide 3 filters, only two of which have a dominant direction. The 1-D bases, from which the 2-D complex bases are derived, are shown to the right.

Test Image and Colour Palette for Complex Coefficients



2-D DT-CWT Decomposition into Subbands

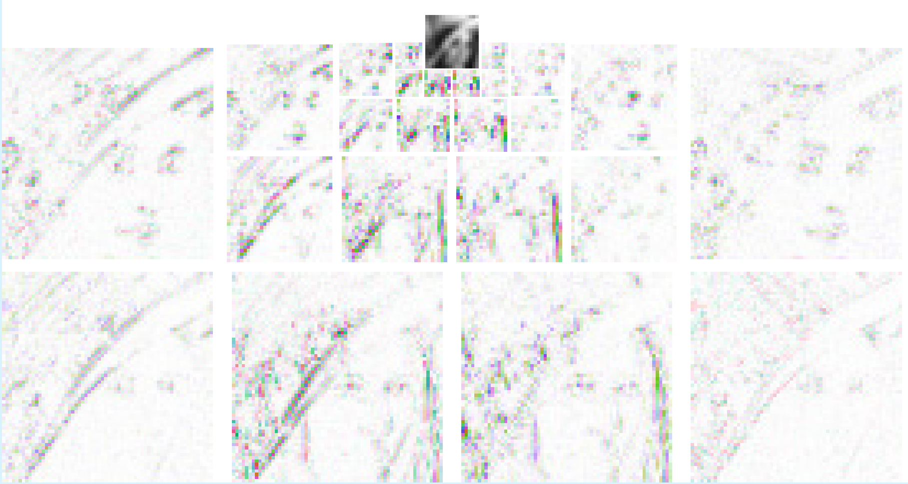


Figure: Four-level DT-CWT decomposition of **Lenna** into 6 subbands per level (only the central 128 × 128 portion of the image is shown for clarity). A colour-wheel palette is used to display the complex wavelet coefficients.

2-D DT-CWT reconstruction components from each subband

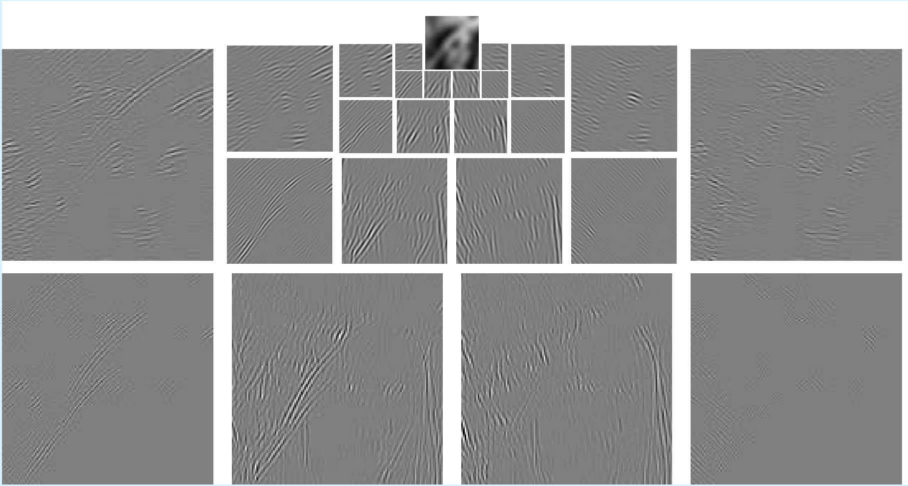
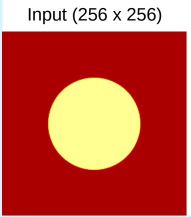


Figure: Components from each subband of the reconstructed output image for a 4-level DT-CWT decomposition of **Lenna** (central 128 × 128 portion only).

2-D Shift Invariance of Complex DT-CWT vs Real DWT



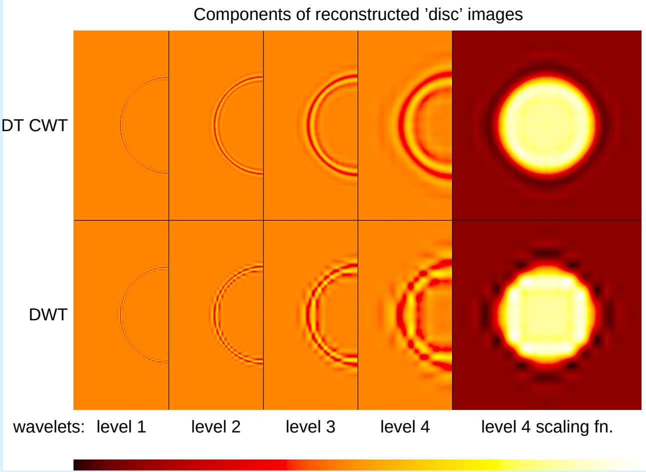
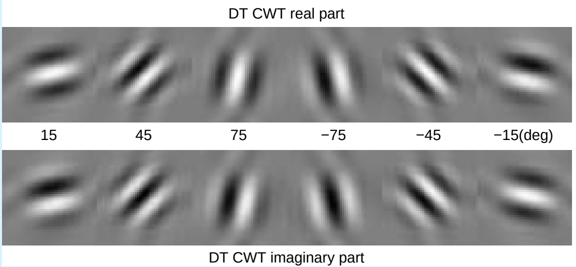
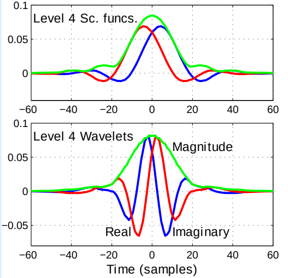
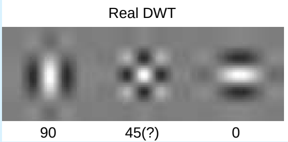


Figure: Wavelet and scaling function components at levels 1 to 4 of an image of a light circular disc on a dark background, using the 2-D DT-CWT (upper row) and 2-D DWT (lower row). Only half of each wavelet image is shown in order to save space.

How do wavelets relate to the human visual system?







$$g(x)\, e^{i\omega_1 x} \,.\, g(y)\, e^{i\omega_2 y} =$$

$$g(x)\, g(y)\; e^{i(\omega_1 x + \omega_2 y)}$$

Compare these basis functions with those found by Hubel and Wiesel in cat brains, and with those found by Olshausen and Field from natural scenes.

Why are dual-tree wavelets good for computer vision systems?

Since 1998 we have used the DT-CWT successfully for the following computer vision tasks:

- **Motion estimation** [Magarey 98]
- **Motion compensation** & **registration** [Kingsbury 02, Hemmendorff 02]
- **Denoising** [Choi 00, Miller 06] and **Deconvolution** [Jalobeanu 00, De Rivaz 01, J Ng 07]
- **Texture analysis** [Hatipoglu 99] and **synthesis** [De Rivaz 00]
- **Segmentation** [De Rivaz 00, Shaffrey 02], **classification** [Romberg 00] and **image retrieval** [Kam & T T Ng 00, Shaffrey 03]
- **Object matching & recognition** [Anderson, Fauqueur & Kingsbury 06]
- **Image fusion** [Nikolov & Bull 07] & **object tracking** [Pang & Nelson 08]
- **Sparse image and 3D-data reconstruction** [Zhang 08 & 10]

Motion Estimation and Tracking Demonstration

- Dual-tree complex wavelet coefficients have the property that they **rotate in phase approximately linearly with displacement / motion** in the direction normal to the stripes of each basis function. We can thus determine motion between frames of a video sequence very efficiently, just by measuring phase shifts between equivalent coefficients from consecutive frames and solving some simple equations.
- Having determined motion, we can then **track objects**, such as people in crowded areas or vehicles at road junctions, and produce many robust motion tracks at once.
- See separate demonstrations.
- We hope to use vehicle tracking as part of a new regional traffic monitoring and congestion prediction system.

Conclusions

- The requirement for sparse representation of natural scenes leads to a dictionary of patches which resemble localised stripes of varying size and orientation.
- The human visual system appears to have evolved to use a very similar set of patches for the early stages of vision in the V1 cortex.
- A simple set of mathematical axioms regarding self-similarity across scale and orthogonality, when applied in 2-D, also result in a set of patches which closely match the above, and which are very efficient to compute using the dual-tree complex wavelet transform.

Hence I believe that the use of the DT-CWT or similar methods, forms a promising avenue for further development of automated vision systems.

Papers on complex wavelets are available at: **www.eng.cam.ac.uk/~ngk/**
A Matlab DT-CWT toolbox is available on request from: **ngk10@cam.ac.uk**