SPARSE REPRESENTATIONS OF IMAGES USING OVERCOMPLETE COMPLEX WAVELETS

Tony Adeyemi and Mike Davies

DSP & Multimedia Group, Queen Mary, University of London, Mile End Road, London E1 4NS

ABSTRACT

This paper describes an algorithm developed for generating sparse representations of images in a quick and efficient way using an overcomplete subband representation. Here we demonstrate the algorithm applied to Kingsbury's popular Dual Tree Complex Wavelet Transform.

1. INTRODUCTION

The Sparse Coding techniques employed in signal processing applications aim to find a suitable configuration within an overcomplete dictionary that concentrates the signal energy in as few non-zero coefficients as possible while maintaining a practical and satisfactory approximation to the original signal. The idea is to exploit the added flexibility available from overcomplete dictionaries to produce highly efficient signal and image representations [1, 2]. In the subband architectures, overcompleteness can also be used to reduce aliasing artifacts when processing an image (e.g. de-noising or deconvolution [3]). However the resulting overcomplete data is a highly redundant representation of the original image. Here we aim to bridge this divide by generating sparse overcomplete subband representations for images based on Kingsbury's popular Dual Tree Complex Wavelet Transform (DT-CWT).

The Dual Tree Complex Wavelet Transform is a useful shift invariant and directionally selective image analysis tool, which is a $4\times$ overcomplete subband decomposition. It also possesses important orthonormal structure that we will exploit to create an efficient sparsity inducing transform. The algorithm presented is a generalization of the Iterative Reweighted Least Squares (IRLS) algorithm and the Jacobi iteration for solving linear equations, and can be interpreted as a Generalized EM algorithm. The Jacobi iteration is used to avoid the expensive matrix inversion necessary in the full IRLS procedure. Surprisingly this introduces minimal reduction in performance.

2. SPARSE DECOMPOSITIONS

We describe $\Phi \in \mathbb{C}^{N\times M}$ as the overcomplete dictionary where $M > N$. The objective is to determine an approximate representation, $\Phi s = x + e$ of a signal x where the coefficients, s, are sparse and we have a small approximation error e.

Our approach to achieving sparsity is to solve the following regularized least squares problem:

$$\hat{s} = \arg\min_s \frac{1}{2v}||x - \Phi s||^2 + \lambda \sum_{k=1}^{M} |s_k|^p \tag{1}$$

where $0 < p \leq 1$, v is the variance of e and λ is a scaling parameter for the sparsity measure. The value of p can be interpreted as controlling the degree of sparsity of the prior placed on s_k. $p = 1$ is equivalent to a Laplacian (L_1) prior on s_k, while for $p \to 0$ the cost function in equation (1) has been identified as placing a Jeffrey's prior on the variance in a Hierarchical Gaussian model [4, 5]. This latter choice is the one we take here (for further discussions of the choice of p see [4]).

3. SPARSE IRLS SOLUTIONS

A simple iterative procedure for minimizing equation (1) is IRLS. Let $W \in \mathbb{R}^{M\times M}$ be a non-negative diagonal weighting matrix. A weighted least squares estimate for s can be obtained through matrix inversion as follows:

$$\hat{s} = \left(W^{-1} + \Phi^H \Phi\right)^{-1} \Phi^H x \tag{2}$$

This solution can then be used to solve (1) by iteratively updating the weighting matrix as a function of a previous estimate for s. For example, for the $p \to 0$ case, we get at the ith iterate a re-weighting step of [4]:

$$W_{k,k}^{(i)} = |s_k^{(i-1)}|^2 \tag{3}$$

This can also be identified as a simple EM (Expectation Maximization) algorithm for hierarchical Gaussian models [5] with the weighted Least Squares solution forming the maximization step while the re-weighting is equivalent to the expectation step.

The drawback with using an IRLS scheme with a high dimensional transform such as the DT-CWT is the expensive $M \times M$ matrix inversions, $(W^{-1} + \Phi^H\Phi)^{-1}$, necessary at each iterate. However in generalized EM theory the maximization step can be replaced by any operation that guarantees to decrease the cost function and increase the likelihood [6]. In [4] we introduced a Fast Iterated Reweighted Sparsifier (FIRSP) based on an Expectation Conditional Maximization (ECM) scheme that required the dictionary to be describable as the union of multiple orthonormal bases. Here we will extend this idea to a FIRSP scheme based upon a Jacobi iteration. It utilizes the specific dictionary properties but does not involve a full matrix inversion making it a quick and efficient algorithm.

4. GENERALIZED IRLS AND THE JACOBI METHOD

The Jacobi method is an iterative algorithm for solving linear equations of the form $As = b$. Expanding this out element wise we get:

$$b_i = \sum_{j=1}^{M} A_{i,j} s_j \tag{4}$$

which can be broken into an iterative method as follows:

$$s_i^{(k)} = \big(b_i - \sum_{j \neq i} A_{i,j} s_j^{(k-1)}\big)/A_{i,i} \tag{5}$$

This is the Jacobi method. It can be noted that the order in which the equations are examined is irrelevant, since the Jacobi method treats them independently. For this reason, the Jacobi method is also known as the *method of simultaneous displacements*, since the updates could theoretically be done simultaneously.

We can apply this to form a partial maxmization step in the IRLS as follows. Recall the M-step is solving the following linear equation:

$$\left(W^{-1} + \Phi^H \Phi\right) s = \Phi^H x \tag{6}$$

if we further assume that the image dictionary has been normalized such that the individual columns, ϕ_i have unit norm, applying the Jacobi iteration gives:

$$s^{(k)} = (W^{-1} + I)^{-1} (\Phi^H x + (\Phi^H \Phi - I) s^{(k-1)}) \tag{7}$$

This can be simplified to the following form:

$$s_n^{(k)} = \left(\frac{W_{n,n}}{1 + W_{n,n}}\right) \times (s_n^{(k-1)} + [\Phi^H e^{(k-1)}]_n) \tag{8}$$

where $e^{(k)} = x - \Phi s^{(k)}$ is the reconstruction error at the kth iterate. Under relatively mild conditions this Jacobi iterate can be shown to be stable. Therefore we can guarantee that after a sufficient number of iterates the likelihood will have increased. In fact in all our experiments we have found that using a single Jacobi iterate per EM step is sufficient.

Closer examination of equation (8) shows that each iteration requires a single application of the synthesis transform, Φ, to calculate the reconstruction error and a single application of the analysis transform Φ^H to map back to the coefficient domain. The remaining computation is scalar and therefore requires order M operations. Thus, as long as we use dictionaries with efficient analysis and reconstruction transforms, the resulting algorithm will be fast (a similar exploitation of fast transforms was used in the linear programming algorithms proposed in [2]).

We note that the performance of the Jacobi iterate within the generalized IRLS algorithm goes against the usual intuition for such algorithms. For example it is well known [7] that the using over-relaxation or alternatively a conjugate gradient algorithm should provide better asymptotic performance in solving the linear equations, yet, in our experience, such algorithms perform worse than the Jacobi based scheme. This is probably due to the fact that we are not trying to fully solve the linear equations and typically use only one iterate of the Jacobi algorithm per EM step.

5. SPARSE DT-CWT DECOMPOSITIONS

The DT-CWT was first introduced by Kingsbury [3] as a means of overcoming the problems of strong shift dependence and poor directional selectivity of the traditional Discrete Wavelet tranform (DWT) when used in image analysis. Shift dependence is due to the fact that the DWT is a real critically sampled subband filterbank. Furthermore, in 2D, separable filtering of the columns and rows of an image produces four sub images at each level. The Lo-Hi and Hi-Lo band-pass sub-images select mainly vertical or horizontal edges respectively, but the Hi-Hi sub-image contains components from diagonal features of either orientation hence providing poor directional selectivity.

Complex wavelets provide shift invariance and good directional selectivity with a reasonable increase in signal redundancy. The four times redundancy obtained is often considered as an acceptable trade off in image analysis. However developing a suitable Complex Wavelet Transform with perfect reconstruction and good filter characteristics is difficult. Kingsbury's Q-Shift DT-CWT provides a good solution to this problem [8] yielding a transform with appealing features for a wide range of signal and image processing applications.

Separable filtering of the columns and rows of an image produces six bandpass bands at each level which are strongly oriented at angles of approximately $\pm 15^o$, $\pm 45^o$ and $\pm 75^o$ and can be used to generate 2D wavelet transforms that are more directionally selective than the separable 2D DWT. Since the filters are complex in nature, in each direction, one of the two wavelets can be interpreted as the real part of a complex-valued 2D wavelet, while the other wavelet can be interpreted as the imaginary part of a complex-valued 2D wavelet. The complex nature of these wavelets provides approximate shift invariance perpendicular to the wavelets orientation. These properties can be seen visually from the images of the wavelets in figure 1.

Alternatively the DT-CWT can be viewed as providing a set of generating functions, shown in figure 1 that provide good descriptions of the basic building blocks of an image: namely multi-resolution, orientable edges. As such the DT-CWT should also be a good candidate for efficient coding of an image, as long as we can select a sparse representation from the overcomplete dictionary.

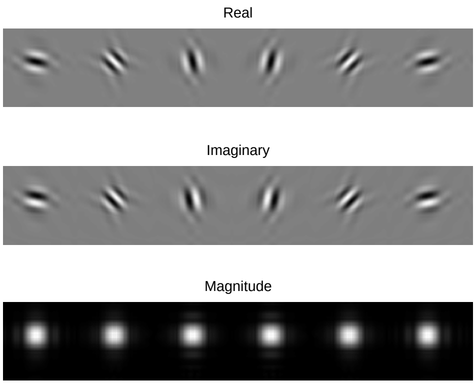

Fig. 1. The real and imaginary components of the level 4 2D DT-CWT wavelets along with their magnitude.

Subsequently we will use Kingsbury's DT-CWT 8 tap filters in the first level and 14 Tap Q-shift for the remaining N levels of the transform [8] using the matlab implementation of Selesnick[1].

With a slight abuse of notation the DT-CWT can be described as the union of four real orthonormal bases. Let $\Phi = (\Phi_{1r}, \Phi_{1i}, \Phi_{2r}, \Phi_{2i})$. with coefficients associated with the real-real, real-imaginary, imaginary-real and imaginary-imaginary parts of the 4 times overcomplete transform. For the 2D DT-CWT, Φ_1 and Φ_2 represent the first and second DT-CWT trees respectively.

The Jacobi iterate described above can most easily applied to the individual orthonormal basis coefficients separately. The re-weighting step then requires the calculation of two diagonal weighting matrices (one for each tree). The re-weighting matrices take the form:

$$W_k = \mathrm{diag}\big(|c_{kr}|^2 + |c_{ki}|^2\big), \text{ for } k = 1, 2 \qquad (9)$$

where c_{1r} are the coefficients associated with the Φ_{1r}, etc.

Given the orthonormal structure for the DT-CWT exposed above one might be tempted to try to use an ECM based algorithm following [9]. However the DT-CWT is not actually a union of orthonormal bases and hence the ECM algorithm fails to work in this instance. In contrast, as shown below the Jacobi algorithm performs very well.

[1]available at `http://taco.poly.edu/WaveletSoftware/`

6. SPARSIFYING AN IMAGE

To demonstrate our approach we apply the sparsfying algorithm to a 256×256 Lena image with 5 levels of the 2-D DTCWT wavelet decomposition. The image was sparsified using a value of $v = 10^{-2}$ and 15 iterations of the algorithm were computed for convergence. Figure 2 shows the coefficients obtained from the standard DT-CWT analysis. Note that the coefficients are already fairly sparse. Applying our iterative algorithm we force the very small coefficients to zero while adapting the remaining non-zero coefficients. This transforms the $4\times$ overcomplete representation into one with only 1.19% of the coefficients nonzero. The resulting approximation had a PSNR of approximately 30dB. The resulting sparse DT-CWT coefficients are shown in Figure 3.

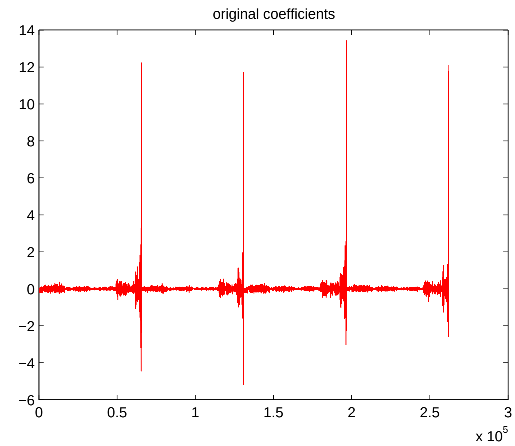

Fig. 2. The 2D DT-CWT coefficients of the Lena image (real and imaginary components concatenated).

Figure 4 is an illustration of the PSNR calculated as $10 \log_{10}(255^2 N/||e_i||^2)$ where e is the error between the input and sparsified image per iterate and N is the number of pixels in the image against the number of non-zero coefficients. The broken curve shows degradation in the PSNR with 100000, 40000, 16000, 8000, 4000, 2000 and 1000 of the non-zero significant coefficients of the DT-CWT. The solid curve with crosses shows the evolution of the PSNR per iteration of the sparsifying algorithm. This shows us how the algorithm preserves the PSNR while aggressively forcing most of the coefficients to zero. To get a visual impression of the sparse approximation, figure 5 shows the original Lena image along with its sparse approximation

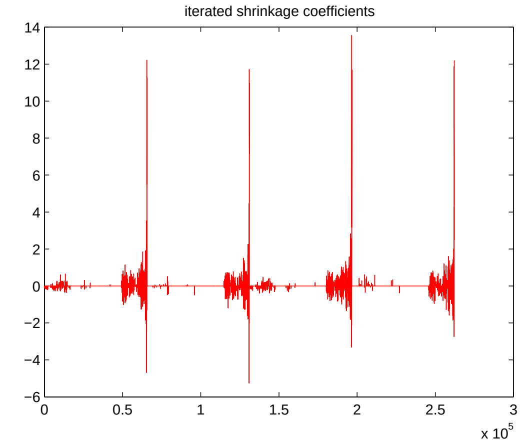

Fig. 3. The sparse 2D DT-CWT coefficients of the Lena image (real and imaginary components concatenated).

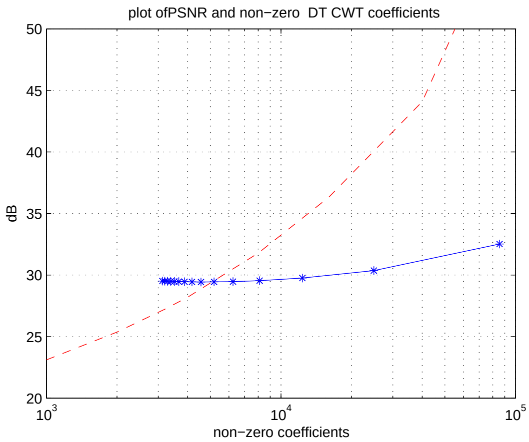

Fig. 4. Plot of PSNR against the number of non-zero coefficients for decreasing values of DT-CWT coefficients and the sparse DT-CWT coefficients per iteration of the algorithm.

which uses less than 3200. While the image quality is degraded compared to the original it is still quite sharp. Compare this with using the 3200 largest wavelet coefficients without sparsification: figure 5(c). Here the reconstruction is visibly blurred in comparison to the figure 5(b).

7. DISCUSSION

A process for generating sparse subband decompositions of images using the DT-CWT has been presented. The algorithm developed shows high speed, low computational cost and typically we require as few as 10 iterations to achieve a good level of sparsity.

In future work we will explore the application of such sparse representations to signal processing and coding. For

(a)

(b)

(c)

Fig. 5. (a) The original Lena image; (b) the sparse reconstructed image, using less than 3200 coefficients and (c) using the largest 3200 non-sparse wavelet coefficents.

example the high degree of sparsity should help in enhancing a noisy image. Also the highly compact description of the image should make this a good candidate for coding despite the redundancy in the original transform. Using the probabilitistic interpretation of the IRLS structure it should also be possible to extend this algorithm to induce structured sparsity by introducing multi-resolution priors.

8. REFERENCES

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[2] S. S. Chen, D. L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[3] N. G. Kingsbury, "Image processing with complex wavelets," *Philosophilcal Transactions of Royal Society of London Series A.*, , no. 357, pp. 2543–2560, 1999.

[4] M. Davies and L. Daudet, "Sparsifying subband decompositions," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to audio and Acoustics*, 2003.

[5] Mario A. T. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1150–1159, Sept. 2003.

[6] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, 1997.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computaions*, John Hopkins, 2 edition, 1989.

[8] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, , no. 10, pp. 234–252, 2001.

[9] M. Davies and L. Daudet, "Fast sparse subband decomposition using firsp," in *Proceedings of EUSIPCO 2004 the 12th European Signal Processing Conference EUSIPCO 2004*, 2004.