IMAGE DENOISING USING OVER-COMPLETE WAVELET REPRESENTATIONS

Slaven Marusic, Guang Deng# and David B. H. Tay#*

*School of Electrical Engineering and Telecommunications, The University of New South Wales
Sydney, N.S.W. 2052, Australia
email: s.marusic@unsw.edu.au
#Department of Electronic Engineering, La Trobe University
Victoria 3086, Australia
email: {D.Deng, D.Tay}@latrobe.edu.au

ABSTRACT

Wavelet transforms have been utilised effectively for image denoising, providing a means to exploit the relationships between coefficients at multiple scales. In this paper, a modified structure is presented that enables the utilisation of an unlimited number of wavelet filters. An alternative denoising technique is thus proposed with a simple approach for the utilisation of multiple wavelet filters. According to the probability distribution function associated with each subband of the transformed data (modelled by generalised Gaussian distribution), different denoising methods are adaptively applied. The proposed expansion is based on the use of either a Walsh-Hadamard Transform (WHT) or independent component analysis (ICA) to remove dependencies between the data streams associated with each wavelet decomposition. The application of a number of different separable wavelet combinations along the rows and columns of the image are also explored.

1. INTRODUCTION

The wavelet transform has been widely shown to be a powerful aid in the removal of Gaussian noise from natural images. Based on the concept of wavelet based image denoising by soft thresholding of wavelet coefficients [1], a number of techniques have been developed to further exploit the dependencies between wavelet coefficients across multiple scales. Among these, the utilization of undecimated wavelet transforms have demonstrated performance improvements in denoising applications while offering useful properties such as shift-invariance. More recently, the development of complex wavelet transforms has demonstrated near shift invariance while reducing the associated computational cost of producing an overcomplete wavelet representation [2]. This dual-tree complex wavelet transform, through careful filter design also produces more directionally selective filters than conventional separable wavelet filters applied in two dimensions. The observed signal (or image) can be modelled as

$$\mathbf{x} = \mathbf{s} + \mathbf{r}$$

where $\mathbf{s}$ is the signal and $\mathbf{r}$ is the noise. This is represented in the wavelet domain as

$$\mathbf{y}_h = \mathbf{w}_h + \mathbf{n}_h \qquad (1)$$

where $\mathbf{w}$ is the original wavelet coefficient, $\mathbf{n}$ is the Gaussian distributed noise and h is an index of the wavelet decompositions. In [3] the respective vectors consist of corresponding parent and child coefficients and MAP estimation is used to estimate $\mathbf{w}$ from $\mathbf{y}$, given by

$$\mathbf{w} = \arg\max_{\mathbf{w}} p(\mathbf{w} \mid \mathbf{y}) \qquad (2)$$

utilising a bivariate distribution to model the relationship between these parent and child coefficients located respectively at different scales of the multiresolution decomposition. A key feature of such denoising techniques is the ability to accurately estimate the local variance of the wavelet coefficients and the associated noise component. Together with the bivariate shrinkage function, a locally adaptive variance estimation of the wavelet coefficients is also applied. In the author's implementation of [3], following the decomposition of the image into multiple data streams by the Dual Tree–Complex Wavelet Transform (DT-CWT), pairs of signals (real and imaginary) are decorrelated using the Walsh-Hadamard Transform (WHT). Following this preprocessing stage, the bivariate shrinkage function is applied, after which the inverse WHT and inverse DT-CWT are applied to reconstruct the images.

In this paper a modified structure that also produces an overcomplete wavelet representation is explored in the context of the image denoising application. The utilisation of separable wavelets for two-dimensional image processing is well documented. In this application, rather than applying the same wavelet in both dimensions, different filters are applied to the rows and columns respectively. This paper also explores the significance of the WHT decorrelation stage and its impact upon the performance of the proposed denoising model. An alternative method for removing statistical dependencies between the different sets of wavelet coefficients is thus proposed, using Independent Component Analysis (ICA). The resulting denoising structure enables, in principle, for the utilisation of an unlimited number of wavelet filters. The proposed structure for the over-complete wavelet representation is presented in Section 2. The signal estimation and denoising model is presented in Section 3.

This work was supported by an Expertise Grant, The Victorian Partnership for Advanced Computing (VPAC) Australia; and the Faculty Research Grant/Early Career Researcher Grant Program, Faculty of Engineering, UNSW, Australia.

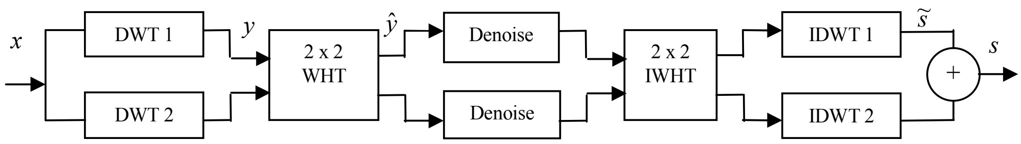

Figure 1. Wavelet based denoising using the Dual-Tree DWT and WHT.

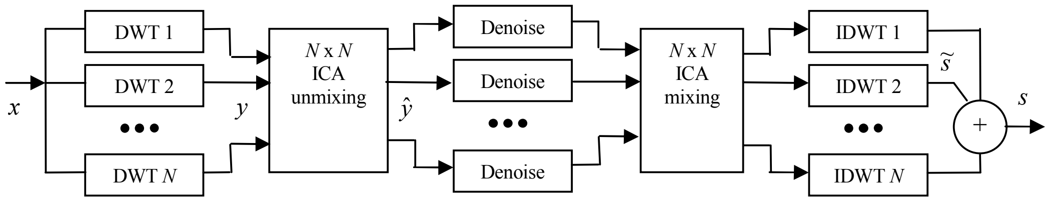

Figure 2. Wavelet based denoising using multiple wavelets and ICA

2. MULTIPLE WAVELET DENOISING (MWD) STRUCTURE

The proposed structure for producing an over-complete wavelet representation for image denoising modifies that utilised in [3] and illustrated in Figure 1 in a number of ways. Firstly, the criteria for the selection of filter pairs are relaxed, as the complex wavelet transform is not applied. Secondly, the model is not restricted to a dual-tree wavelet transform. Thus, where Figure 1, represents DWT1 and DWT2 as a pair of real and imaginary filters, the proposed model allows for the expansion of the number of filters to any number of different filter pairs, or even individual filters. Intuitively, this could enable a larger number of different image characteristics to be captured by the appropriate filters. So, while denoising (in [3]) is performed on the magnitude of the real and imaginary components of the DT-CWT, the MWD structure combines the respective data streams through either the pairwise WHT or through ICA. While sacrificing the approximate shift-invariance of the DT-CWT, the MWD structure is able to produce comparable denoising results.

In utilising a separable filter pair approach, the existing approach can be extended by simultaneously applying all combinations of a given filter pair in both directions. Therefore, from one pair of filters, four different representations of the noisy image are produced (DWT1 applied to rows and columns; DWT2 applied to rows and columns; DWT1 applied to rows with DWT2 applied to columns, and vice versa). This simple extension produces consistent improvements in PSNR compared with limited selections of filter orientations. The 2x2 WHT is applied to the output of these DWTs, combining $y_{DWT1,DWT1}$ with $y_{DWT2,DWT2}$ and combining $y_{DWT1,DWT2}$ with $y_{DWT2,DWT1}$, where the subscript represents the filters used along rows and columns respectively.

2.1 MWD using Independent Component Analysis

Independent Component Analysis (ICA) is a signal processing tool for extracting independent data structures from complex data sets [4]. Commonly used in blind source separation applications, where statistically independent sources have been linearly mixed and observed by multiple sensors, ICA is used to separate the original sources without any knowledge of the mixture. Using conventional ICA notation, an observed signal $\mathbf{x}$ is given by

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

where $\mathbf{A}$ is the mixing matrix applied to the independent signal channels ($\mathbf{s}$) and the output is given by

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

and $\mathbf{W}$ is the unmixing matrix. Given the assumption of statistically independent sources, ICA utilises the principle of maximisation of mutual information based on the joint probability density functions of the mixed sensor data to perform the separation. Compared with common decorrelation techniques, such as PCA that only remove second order dependencies, ICA exploits higher order statistical relationships.

Rather than applying the WHT to decorrelate the signal pairs and having observed the inferior performance of the 4x4 WHT, an alternative structure utilising ICA is proposed that offers greater flexibility in the number image representations used for denoising (Figure 2). In this application, a number of different DWT's are applied to the input image producing multiple transformed versions of the original noisy image. These two-dimensional data sets are then converted to one-dimensional signals such that the relative scales are identically ordered between the respective signals. The statistical redundancy between the data streams is reduced by first calculating and then applying the ICA unmixing matrix to the data. These independent data streams are then individually denoised using a given denoising technique. The process is then inverted to recover the now denoised images, first by applying the ICA mixing matrix and then the respective inverse DWT's. These are then averaged together to produce the final denoised version of the original noisy image.

The proposed denoising model allows for the inclusion of additional wavelet decompositions of the original image. The outcome is the ability to utilize greater combinations of wavelets to effectively capture a wider range of image characteristics. In particular, this approach enables the application of different wavelets along rows and columns respectively while incorporating all of the possible orientations from the selected wavelets. The proposed model based on the ICA mixing matrix has been observed to produce results superior to alternative implementations employing 4 x 4 WHT's. Furthermore, this structure is not restricted to the pairwise DWT approach described in Section 2, and thus allows for an odd number of image representations.

3. SIGNAL ESTIMATION AND DENOISING

The overcomplete wavelet representations described thus far provide an effective means by which to enhance the performance of image denoising techniques. However, the key to effective denoising based on the signal decomposition given by (1) remains the ability to accurately separate the original signal from noise.

The denoising problem can be divided into two sub-problems: estimating w_k from its noise-corrupted observation y_k and estimating the original signal s_k from a set of denoised wavelet coefficients w_k. Both problems can be formulated as a MAP estimation problem. For example, the second problem can be described as

$$s = \arg\max_{s} p(s \mid \tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_M)$$

where s_k is the inverse wavelet transform of w_k and $\tilde{s}_h$ is the estimate of the signal from a given channel. To simplify our discussion, we assume the following model:

$$s = \tilde{s}_h + e_h$$

where e_h for h = 1,2,…,M are iid zero-mean Gaussian. With this assumption, it can be easily shown that

$$s = \frac{1}{M}\sum_{h=1}^{M} \tilde{s}_h$$

The first sub-problem can be expressed as

$$\mathbf{w} = \arg\max_{\mathbf{w}} p(\mathbf{w} \mid \mathbf{y})$$

where $\mathbf{w} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M\}$ and $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$. To simplify the estimation process, we need to decorrelate the elements of $\mathbf{w}$. If the elements of $\mathbf{w}$ are modelled by a joint Gaussian distribution then the WHT can be used. On the other hand, if a generalised Gaussian distribution is assumed, then ICA is a suitable candidate for the task. Applying one of these 'decorrelation' techniques we obtain:

$$p(\hat{\mathbf{w}} \mid \hat{\mathbf{y}}) \propto p(\hat{\mathbf{w}}_1 \mid \hat{\mathbf{y}}_1) p(\hat{\mathbf{w}}_2 \mid \hat{\mathbf{y}}_2) \ldots p(\hat{\mathbf{w}}_M \mid \hat{\mathbf{y}}_M) \quad (3)$$

where $\hat{g}$ is the decorrelated version of g. Taking the log of (3) we obtain,

$$\log p(\hat{\mathbf{w}} \mid \hat{\mathbf{y}}) = \log p(\hat{\mathbf{w}}_1 \mid \hat{\mathbf{y}}_1) + \log p(\hat{\mathbf{w}}_2 \mid \hat{\mathbf{y}}_2) + \log p(\hat{\mathbf{w}}_M \mid \hat{\mathbf{y}}_M)$$

The aim is thus to effectively decouple the signal from the individually transformed observations to enable each channel to be individually denoised.

Taking a generalised Gaussian function

$$\mathrm{p}(x) \propto a e^{-(\alpha|x|)^{\beta}}$$

and then determining the parameters α and β from each subband, the given distribution function can be efficiently estimated [7] allowing for the appropriate denoising strategy to be selected. It is well established that $\beta = 1$ gives a Laplacian distribution for which soft thresholding has been demonstrated as an effective denoising approach. Similarly, shrinkage denoising is suitable for a Gaussian distribution ($\beta = 2$). Improved results should therefore be expected by not only applying appropriate denoising to these specific cases but also for $0<\beta<1$, where the shrinkage function tends towards a hard threshold nonlinearity [8].

Using ICA as a decoupling stage can produce a variety of such outputs, thus placing greater emphasis specific denoising functions to be applied to the respective channels or subbands. The application of ICA therefore does not restrict the denoising structure to the assumption of joint Gaussian relationships between the filter pairs, nor does it limit the structure to pairwise representations of the observed noisy data. The consequence of this being the potential for an increased number of image representations (or wavelet decompositions) each tailored to particular image characteristics.

After we decouple the channels, the objective then becomes to estimate the signal in each individual channel. This is performed in conjunction with a suitable denoising technique. If the transformed noisy observations are locally modelled by a joint Gaussian distribution, the WHT is utilised as a decoupling stage. In this case M = 2, such that pairwise WHT decompositions are employed. Producing Gaussian outputs, a joint Gaussian based shrinkage function is then a suitable approach for denoising.

Applied here is an MAP estimator (2) to perform the image denoising, such that the signal estimate is given by

$$w_k = \frac{\sigma_w^2}{\sigma_w^2 + \sqrt{2}\sigma_n^2} y_k \quad (4)$$

and assuming both noise and signal components are Gaussian distributed, the estimation error for a Wiener filter is

$$\sigma_e^2 = \frac{\sigma_w^2 \sigma_n^2}{\sigma_w^2 + \sigma_n^2}$$

The noise variance is obtained with a robust estimator applied to the high pass wavelet coefficients at the highest resolution (subband HH_1), given by:

$$\sigma_n = \mathrm{median}(|\mathbf{y}|)/0.6745$$

The wavelet transformed signal variance is given as follows,

$$\sigma_w^2 = \begin{cases} E_k - \sigma_n^2, & E_k > \sigma_n^2 \\ 0 & otherwise \end{cases}$$

The signal estimate E_k is determined for each coefficient based on the signal energy calculated from a local two dimensional area as follows

$$E_k = \frac{1}{(2\mathrm{N}+1)^2} \sum_{i=N-1}^{N} \sum_{j=N-1}^{N} y_m(k-i, k-j)^2$$

where $y_m(u,v)$ represents the wavelet coefficient at position (u,v) at scale m. This can be further refined by linearly combining the estimates from two differently sized areas, so that

$$E_k = \sum_{r=1}^{R} \alpha_r E_{k,N(r)} \quad (5)$$

Alternatively, if it is assumed that the transformed noisy observations in fact produce outputs with generalised Gaussian distributions then a different decoupling stage and associated denoising method may be more appropriate. From a localised viewpoint, the wavelet coefficients are still modelled as Gaussians (given that non-Gaussian distributions can in turn be modelled as a mixture of Gaussians). However, by estimating β for each subband [7], the signal estimate (4) can be modified such that the shrinkage function tends more to a hard threshold for $0<\beta<1$, by incorporating β as follows,

$$w_k = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_n^2 \beta\sqrt{2}} y_k \qquad (6)$$

4. RESULTS AND DISCUSSION

Presented in Table 1 are a number of results which demonstrate the effectiveness of the proposed MWD denoising techniques applied to noisy versions of the Barbara image. The results from the proposed techniques are averaged over 50 simulations to smooth the effect of the random noise distributions applied to the original image. The proposed techniques are contrasted with results reported from some of the existing leading techniques in the literature (with the respective results given as reported in [3]. Independent Component Analysis is implemented using FastICA [4].

The effectiveness of the multiple wavelet denoising (MWD) structure is demonstrated by comparing MWD-WHT (A) with (B). Version (A) applies the filters 'sym12' and 'coif2' (see MATLAB) in both directions, respectively to produce two wavelet decompositions. Version (B) applies all combinations of the same filter pair along rows and columns to produce four decompositions. The improvement achieved by channel decoupling is also shown by comparing MWD with MWD-WHT or MWD-ICA.

Investigations involving other filter combinations reveal that even when utilising the WHT (or ICA) stage, careful selection of wavelet filters is still required, as inappropriate combinations can degrade the denoising performance. Similarly, increasing the number wavelets can further improve performance, evident from the results of MWD-ICA (C), which utilises 'sym12', 'coif2' and 'db7' to produce three channels. Further increasing the number of channels, by applying all pairwise combinations of the three wavelets along rows and columns respectively (refer MWD-ICA (D)), continues to improve the performance.

Incorporation of the generalised Gaussian shape parameter β in the signal estimation (or denoising) function, exploits the suitability of hard thresholding for subbands where $0<\beta<1$. The effectiveness of the method proposed in (6), particularly for high noise variance, is demostrated in the results of MWD-ICA-GSP (C) and (D). A subband specific densoising function is thus applied to exploit the variation in distributions observed across all channels and all subbands.

5. SUMMARY

In this paper, we have explored the effectiveness in image denoising of an alternative multiple wavelet denoising structure together with a simple adaptive denoising technique. The proposed denoising structure was demonstrated to provide performance improvements and comparable results to existing leading techniques. Additional flexibility in the number of wavelet decompositions to be applied was also demonstrated through the use of ICA based channel decoupling.

REFERENCES

[1] D. L. Donoho, "De-noising by soft thresholding," *IEEE Trans on Information Theory*, 41, pp. 613-627, 1995.
[2] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, Vol. 10, No. 3, pp. 234-253, May 2001.
[3] L. Sendur and I. W. Selesnick, "Bivariate Shrinkage with Local Variance Estimation," *IEEE Signal Processing Letters,* Vol 9. No. 12, pp. 438-441, December 2002.
[4] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, 13(4-5):411-430, 2000.
[5] M. K. Mihcak, I. Kozintsev, K. Ramchandran and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, Vol. 6, pp. 300-303, Dec. 1999.
[6] X. Li and M. T. Orchard, "Spatially adaptive image denoising under over-complete expansion," in *Proc IEEE International Conference on Image Processing*, Sept. 2000
[7] K. Sharifi and A. Leon-Garcia, "Estimation of Shape Parameter for Generalized Gaussian Distriutions in Subband Decompositions of Video," *IEEE Trans. Circuits and Systems for Video Tech.*, Vol. 5, No. 1, pp. 52–56, Feb 1995.
[8] P. Moulin and J. Liu, "Analysis of Multiresolution Image Denoising Schemes Using Generalized Gaussian and Complexity Priors," *IEEE Trans. on Information Theory*, Vol. 45, No. 3, pp. 909-919, April 1999.

Table 1. Denoising results – PSNR (dB).
A: {sym12, coif2}; B: {A (all combinations)}; C: {A,db7}; D: {C (all combinations)};

Noise variance	LAWMAP [5] (ref. [3])	[6] (ref. [3])	Bivariate shrinkage [3]	MWD (A)	MWD – WHT (A)	MWD – WHT (B)	MWD – ICA (B)	MWD – ICA (C)	MWD – ICA (D)	MWD – ICA – GSP(C)	MWD – ICA – GSP(D)
10	31.99	33.35	33.35	33.41	33.51	33.66	33.59	33.57	33.63	33.58	33.65
15	29.60	31.10	31.31	31.07	31.18	31.35	31.29	31.25	31.27	31.28	31.33
20	27.94	29.44	29.80	29.41	29.51	29.70	29.59	29.60	29.57	29.67	29.69
25	26.75	28.23	28.61	28.15	28.20	28.42	28.27	28.32	28.26	28.44	28.44