Image denoising based on adaptive spatial segmentation and multi-scale correlation in directionlet domain

R. Sethunadh*
Vikram Sarabhai Space Centre, Indian Space Research Organisation (ISRO), Trivandrum, Kerala, India
Email: r_sethunadh@vssc.gov.in
*Corresponding author

Tessamma Thomas
Image and Video Processing Laboratory, Cochin University of Science and Technology (CUSAT), Cochin, Kerala, India
Email: tess@cusat.ac.in

Abstract: Directionlet transform (DT) has become popular over the last few years as an efficient image representation tool due to its fine frequency tiling and directional vanishing moments along any two directions. A novel denoising algorithm based on DT is proposed here for images corrupted with Gaussian noise. The image is first spatially segmented based on the content directionality. Then an undecimated version of DT is applied to effectively capture the directional features and edge information of these segments. The DT coefficients so obtained are then modelled using a bivariate heavy tailed Gaussian distribution and the noise free coefficients are computed using MAP estimator. By employing bivariate probability distribution, the heavy-tail behaviour of natural images is accurately modelled and the interscale properties of DT coefficients are properly exploited. In addition, the local variance parameter of the model is estimated based on classification of DT coefficients within a particular scale using context modelling. Due to this the intrascale dependency of the directionlet coefficients is also well exploited in the enhancement process. The proposed algorithm is competitive with the existing algorithms with better results in terms of output peak signal-to-noise ratio while having comparable computational complexity. It exhibits good capability to preserve edges, contours and textures especially in images with abundant high frequency contents.

Keywords: directionlet transform; image denoising; multi-scale correlation; spatial segmentation; interscale dependency.

1 Introduction

Separable two-dimensional (2-D) wavelets have been one of the major research tools for image representation over Fourier basis. Over the past two decades many image denoising schemes based on wavelet transform (WT) (Daubechies, 1990) were proposed. These methods can be broadly classified into two: threshold-based methods and statistical model-based methods. In threshold-based methods, a WT coefficient is compared with a given threshold and is set to zero if its magnitude is less than the threshold; otherwise it is kept unmodified or modified depending on hard or soft thresholding rules, respectively. The effectiveness of these methods depends on the estimation of the correct threshold. Of the various thresholding strategies, soft-thresholding is the most popular one and has been theoretically justified by Donoho (1995). The model-based methods are based on statistical modelling of WT coefficients with prior probability distribution functions. The noise free coefficients are then estimated using this a priori information with Bayesian estimation techniques, such as the maximum a posteriori (MAP) estimator. Here the main problem is effectively modelling the image and noise coefficients. If these models are well chosen, the noise can be removed efficiently.

It was well established that human eye perceives the surrounding world in geometrical multi resolution way (Olshausen and Field, 1996). Thus an efficient image representation scheme should be geometrical and multiresolutional.

The commonly used discrete WT (DWT) is not geometrical and is thus unsuited to exploit the correlation along edges and contours in images. Various multi-scale geometrical transforms with directional selectivity have been developed for image representations over the past decade as generalisations of the wavelet theory. These include adaptive or signal dependent schemes and non-adaptive schemes. The non-adaptive methods are characterised by fast transforms and are based on frames. The examples of such schemes are curvelet (Candès and Donoho, 1999), contourlet (Do and Vetterli, 2005), shearlets (Kutyniok and Labate, 2009), dual tree complex WT (DTCWT)
Curvelet is defined as a tight frame used to represent smooth objects having discontinuities along smooth curves. Contourlets are defined as a discrete version of curvelets based on a double filter bank structure by combining the Laplacian pyramid with a directional filter bank (Bamberger and Smith, 1992). The basis functions of these transforms have wedge shaped or rectangular support regions which provide good sparse representations for high dimensional singularities. Shearlets are generated by parabolic scaling, shearing and translation operations. Even though many image processing applications based on these transforms are available, the sub sampling operations involved in the multi-scale partition and directional filter processing of these transforms cause pseudo-Gibbs phenomenon and lack of shift invariance which adversely affect the performance. One other major drawback of these transforms is the higher computational complexity as compared to separable 2-D WT. The DTCWT is an over complete WT, which is implemented by two wavelet filter banks operating in parallel. It exhibits approximate shift invariant property and better directional selectivity in multiple directions with reduced computational complexity. The DTCWT results in an approximation subband and six directional subbands at each level, which are strongly oriented at angles of $\pm 15^\circ$, $\pm 45^\circ$ and $\pm 75^\circ$. The main drawback with the above mentioned non-adaptive schemes is that they use only prefixed standard directions instead of the actual dominant directions in the image to decompose it resulting in an inefficient sparse representation.

The adaptive methods of image representation are based on dictionaries or bases. The best examples of adaptive schemes based on dictionaries are wedgelets (Donoho, 1999), beamlets (Donoho and Hou, 2000) and smoothlets (Lisowska, 2011). These schemes are known to be more efficient than non-adaptive ones, since a dictionary can be defined more accurately than a frame. The main disadvantage of these schemes is the high computational complexity. Also since these schemes are defined on discontinuous functions, they can efficiently represent only the well defined edges in images. The methods based on bases are usually implemented in a multi resolution filter bank way and are relatively fast. The best known adaptive schemes based on bases are bandelets (Le Pennec and Mallat, 2005), brushlets (Meyer and Coifman, 1997), grouplets (Mallat, 2009), tetrolets (Krommweh, 2009), directionlet (Velisavljević et al., 2006) and WT-based schemes like orientation adaptive WT (Taubman and Zakhov, 1994), directional lifting WT (Ding et al., 2007; Chang and Girod, 2007) and curved WT (Wang et al., 2006). Bandelets are constructed as an orthonormal basis for the approximation of geometrical boundaries whereas brushlets are defined as an adaptive basis of functions, which are well localised in only one peak in frequency. Grouplets use orthogonal bases which are defined to group pixels to represent geometrical regularities. Tetrolets are based on adaptive Haar WT performed on specific domains of tetromino shapes. The directional WT and curved WT are based on directional lifting and keep the down sampling pattern same as that of standard WT, i.e., vertical down sampling followed by horizontal down sampling or vice-versa, and vary the filtering direction locally. However, due to the possible mismatch between the down sampling and the filtering directions, these transforms may suffer from aliasing. On the other hand, orientation adaptive WT, and directionlets apply both filtering and down sampling along the dominant directions. The orientation adaptive WT allows filtering and down sampling along any two arbitrary directions using an invertible re-sampling involving interpolation of pixels at arbitrary locations, whereas, directionlets allow filtering and down sampling along any two arbitrary rational directions by applying 1-D WT along the lines defined...
Image denoising based on adaptive spatial segmentation

on integer lattices without any interpolation. Both these conceptually similar methods apply spatially varying re-sampling followed by separable filtering, and hence, are forced to process on segmented image (Jayachandra and Makur, 2014).

Owing to the fact that multi-scale geometric transforms provide image representations of high-energy concentration, the image denoising methods based on these transforms generally outperform DWT-based methods. Image denoising imposes a compromise between noise reduction and preserving significant image details. To achieve a good performance in this respect, a denoising algorithm has to adapt to image discontinuities. Geometrical features in images, like edges and contours, play one of the most important roles in the human visual system, since they carry most of the perceptual information. An efficient image representation has to be capable of precise modelling and of providing a sparse description of this geometrical information. The directionlets has been proved to provide sparse representation of images and to improve the performance of wavelet-based image processing methods. Recent works on directionlets in different image processing problems like despeckling (Sethunadh and Thomas, 2013), super resolution (Reji and Thomas, 2011), image compression (Velisavljević et al., 2007), etc., show its superiority over DWT. These achievements motivate us to use directionlets to improve the available image denoising schemes in terms of edge and feature preservation.

The effective modelling of the statistics of signal and noise plays a major role in the performance of statistical model-based denoising schemes. If these models are well chosen, the noise can be efficiently removed. In literature several models have been considered for the noise-free wavelet coefficients and Gaussian model for the noise coefficients. In most of these models the WT coefficients are assumed to be independent. However it was well established that there are inter and intra scale statistical dependency in wavelet coefficients of natural images because if a WT coefficient has small magnitude the adjacent coefficients are very likely to be small, and the small coefficients tend to propagate across the scales. Thus the models which consider the WT coefficient as independent cannot efficiently model the transform coefficients of natural images and thus may not provide good denoising performance. The performance of denoising schemes based on multi-resolution analysis would be significantly improved if the multiscale correlation among the transform coefficients is also taken into account. Theoretically this is true for any transform with multi resolution representation for images. Sendur and Selesnick (2002a) developed a subband adaptive bivariate shrinkage function for image denoising in the wavelet domain using interscale dependency. Later they modified the bivariate shrinkage function with local variance estimation and applied it to the magnitude of the DTCWT coefficients and reported better results (Sendur and Selesnick, 2002b). Chen et al. (2012) proposed a denoising scheme in which the statistical dependency between a DTCWT coefficient and its parent and children across three scales is exploited. Even if DTCWT provides better directional selectivity, it is not adaptive as it uses only prefixed standard directions instead of the actual dominant directions in the image to decompose it resulting in an inefficient sparse representation. In this paper a spatially adaptive image denoising scheme is proposed by considering the multi scale dependency of the DT coefficients.

This paper is organised as follows. Section 2 gives theoretical concepts of DT. In Section 3, the proposed adaptive spatial segmentation and the bivariate shrinkage scheme
in directionlet domain are presented. Section 4 reports the experimental results and finally, conclusions are given in Section 5.

2 Directionlet transform

The DT was proposed as an alternative to 2-D WT to overcome the WT’s main drawbacks of limited directionality and isotropic nature. The DT is an anisotropic WT with directional vanishing moments along any two directions. The isotropic nature of 2-D WT causes a square shaped frequency tiling leading to a large number of wavelets to represent the discontinuities in images. In case of DT, due to its anisotropic nature, the frequency tiling is rectangular in shape resulting in a small number of wavelets to represent the discontinuities as shown in Figure 1. This will lead to many large magnitude coefficients in transform domain in case of WT as compared to DT for the same image.

Figure 1 Representation efficiency of isotropic and anisotropic basis functions, (a) spatial frequency tiling of standard 2-D WT for two iteration and (b) its basis functions (c) spatial frequency tiling of DT with anisotropic ratio of 1:2 for two iteration and (d) its basis functions

The DT is anisotropic and it applies 1-D WT along any two rational directions with slopes \( n_1 = \left( \frac{b_1}{a_1} \right) \) and \( n_2 = \left( \frac{b_2}{a_2} \right) \), where \( a_i \) and \( b_i \) are integers. These two directions can be represented in vector form as \( d_1 = [a_1, b_1] \) and \( d_2 = [a_2, b_2] \). The anisotropy is built in by carrying out unequal number of transforms along the two directions and this will continue in the low pass sub-band for multi resolution expansion. Let \( n_2 \) is the number of
transforms along direction $d_1$ and $n_2$ that along direction $d_2$. In every scale, the DT decomposes an image into $(2^{n_1+n_2} - 1)$ detail subbands with $(2^{n_1+n_2} - 1)$ number of orientations and one approximation subband, which are anisotropic with ratio $n_2: n_1$. The corresponding basis functions in the space domain will have an aspect ratio of $n_1: n_2$ in the horizontal-to-vertical directions. This scheme is iterated on the approximation subband to obtain coarser subbands and can be implemented by a 2-D separable filter bank.

The main problem with this type of transform is the directional interaction due to the possible non-orthogonal transform directions and its anisotropic nature. To overcome this, a lattice-based implementation is recommended. In lattice-based transform the discrete space is first partitioned using integer lattices before performing 1-D filtering along lines across the lattice. Any integer lattice $\Lambda$ can be represented by a non-unique generator matrix $M_{\Lambda}$.

$$M_{\Lambda} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$ (1)

where $a_1, b_1, a_2, b_2 \in \mathbb{Z}$ and $d_1$ and $d_2$ are two linearly independent integer vectors. The linear combination of these two vectors will form the points of the lattice $\Lambda$. The integer lattice $\Lambda$ is a sub-lattice of discrete space $\mathbb{Z}$ which can be partitioned into $|M_{\Lambda}|$ cosets, where each coset is determined by the shift vector, $S_k \in \mathbb{Z}$, for $k = 0, 1, 2, \ldots |M_{\Lambda}| - 1$. The integer lattice $\Lambda$ and its cosets, partitions each digital line with slopes $r_1$ and $r_2$ into co-lines $CL_{\Lambda}(r_1, n)$ and $CL_{\Lambda}(r_2, n)$ respectively (Figure 2). The filtering and sub sampling operations are applied on the pixels in each of the cosets separately. Note that each filtering operation is purely 1-D. Applying the 1-D filtering and sub sampling on the co-lines $CL_{\Lambda}(r_1, n)$ will not disturb that along $CL_{\Lambda}(r_2, n)$, means the retained pixels after these operations belong to the same co-lines as they were before. This way the directional interaction is avoided.

**Figure 2** The intersections between the three cosets of the lattice $\Lambda$ given by the generator matrix $M_{\Lambda}$ and the digital lines with slope $n = 1/2$ (see online version for colours)!  

Note: The co-lines are $CL_{\Lambda}(0,0)(\frac{1}{2},n)$, $CL_{\Lambda}(0,1)(\frac{1}{2},n)$ and $CL_{\Lambda}(1,1)(\frac{1}{2},n)$. 
The directionlet transform is not shift invariant because of the decimation operation involved in the transform construction. In image denoising, a small shift in the input signal can cause very different output wavelet coefficients. One way of overcoming this is by taking the transform without decimation. The drawback of this approach is that it is computationally inefficient, especially in multiple dimensions. In the multi resolution analysis, the oversampled transforms have been shown to lead to better denoising performance than critically sampled transforms. For that reason, the denoising algorithm we propose in this paper is based on the undecimated directionlet transform. Oversampling is imposed in the same way as in the case of the undecimated WT, that is, by discarding the sub-samplers in the filtering process. The LP filters used for DT are up sampled across scales. Thus, if the LP filter \( H_2(Z_1) \) is used in first level 1-D wavelet filter-bank, the equivalent LP filter at the \( j \)-th scale is \( 2^j H_2(Z_1) \), where \( j = 1, \ldots, J \), corresponds to the scale index sweeping from the finest to the coarsest scale. Such a construction results in a shift-invariant transform with a preserved number of coefficients in each subband.

The computational complexity of DT is similar to that of WT which is in the order of \( N^2 L \). Each filtering operation is performed in \( N L \) multiplications and \( N L \) additions. Here \( N \) is the number of input samples and \( L \) is the length of the applied filter. The extra computation is due to the anisotropic nature of DT. The computational complexity is substantially lower than other similar schemes like bandlets, which require \( N^2 (\log_2 N)^2 \) operations and contourlets, which require \( L_1 L_2 N^2 \) operations, where \( L_1 \) and \( L_2 \) is the size of the 2-D filters. However the true benefit of DT comes from the identification of local dominant directions in an image by spatial segmentation and decomposition of these spatial segments using the anisotropic transform along the dominant direction. Thus for the better identification of directional features using DT, extra computations are required for segmentation and content direction identification. This adds to the computational complexity of DT-based applications. So there is a trade off between the computational efficiency and denoising performance in DT-based schemes.

3 Denoising in directionlet domain

3.1 Adaptive spatial segmentation and decomposition

By adapting the transform directions to the local content directions as close as possible, the DT concentrates the signal information to the low frequency approximation sub band as much as possible. However the DT can have directional vanishing moments only along two directions with rational slopes. Thus to get the best benefit out of DT, the dominant directions in an image have to be identified in advance to select the transform directions.

The content directionality in an image varies over space. Therefore it is ideal to find out the directionality in an image locally after spatially segmenting it into small blocks. Here a segmentation scheme adaptive to the content directionality is proposed. First, the input image is divided into four equally sized segments in a step of the quad-tree spatial segmentation. Then for each spatial segment a parameter called directional variance (Jayachandra and Makur, 2010) is computed along eight different directions. The directional variance, \( \text{DirVar}(X, r) \) measures the normalised sum of variances along each
digital line with the given slope \( r \) and hence is very sensitive to content directionality. It is computed as

\[
\text{DirVar}(X, r) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{k_i} (X_j - X_{M(r, i)})^2
\]

where \( X_{M(r, i)} \) is the mean of the digital line with slope \( r \) and offset \( i \), and \( X_j \) is the pixel in the same line. \( N \) is the total number of pixels in the segment \( X \), \( n \) is the total number of lines, and \( k_i \) is the number of pixels in line \( i \). Here we have computed the directional variance along eight different directions with \((i, j) = (1, 0), (2, 1), (1, 1), (1, 2), (0, 1), (-1, 2), (-1, 1) \) and \((-2,1)\) which corresponds to 0º, 30º, 45º, 60º, 90º, 120º, 135º and 150º. Even though more directions are possible, in practice, we have observed that these eight directions are enough to achieve a good performance.

For segments with uniform region or texture which does not have specific directionality, \( \text{DirVar} \) does not vary much from direction to direction. This particular behaviour of \( \text{DirVar} \) enables us to differentiate the uniform and texture regions from the regions with specific directionality. If the value of \( \text{DirVar} \) along all eight directions are within 5% then that segment need no further division. The segments with more than two dominant directions are further quad tree segmented and the process is repeated till the predetermined maximal segmentation depth is reached. The two dominant directions of each segment are then selected as the transform and alignment directions. This way the transform directions are adapted independently in each spatial segment allowing for more efficient capturing of geometrical information. For segments having no apparent dominant directions, the transform directions are selected as (0º, 90º) for simplicity of implantation of DT. The optimal segmentation and the choice of transform and alignment directions for a simulated circle image is shown in Figure 3.

Figure 3  The optimal segmentation and choice of transform directions in each segment identified by directional variance

As mentioned earlier, the directionlets apply transform separately on each coset of the integer lattice given by the chosen generating matrix, and by definition no interaction is allowed between the cosets. Here based on the computational efficiency point of view, we have selected the generating matrices for the DT decomposition in such a way that the determinant of the generating matrices value is one so that it will not form more than one
coset. To satisfy this condition we consider the direction vectors \( d_1 \) and \( d_2 \) such that \(|\det(M_A)| = 1\). One useful subset of matrices with \(|\det(M_A)| = 1\) are of the form

\[
M_A = \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix} \quad \text{and} \quad M_A = \begin{bmatrix} 0 & 1 \\ 1 & a \end{bmatrix}
\]

where \( a \in \mathbb{Z} \). Varying \( a \) leads to a large number of transform directions. Based on the eight directions considered earlier for \( \text{DirVar} \) computation, we can form 13 generating matrices as follows which will satisfy the condition, \(|\det(M_A)| = 1\).

\[
\begin{align*}
M_{A1} & = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_{A2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad M_{A3} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \\
M_{A4} & = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad M_{A5} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}, \quad M_{A6} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \\
M_{A7} & = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}, \quad M_{A11} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad M_{A13} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \\
M_{A18} & = \begin{bmatrix} -1 & 2 \\ 1 & 1 \end{bmatrix}, \quad M_{A19} = \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix}, \quad M_{A21} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \\
M_{A23} & = \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix}
\end{align*}
\]

Using these generator matrices the DT decomposition was carried out for each spatial segment along the content directionality. Pairs of directions that would result in \(|\det(M_A)| > 1\) and thereby form multiple cosets are not considered here to avoid the division of spatially adjacent pixels into different cosets. At every level of the transform only one level of 1-D transform is applied along each direction to avoid computational complexity. This will generate an isotropic decomposition. Even though anisotropic nature provides a faster asymptotic approximation rate it will result in a small number of scales and, therefore, the multi-scale dependence of the coefficients cannot be properly captured. Symmetric extension at the borders of the segment was carried out while taking the transform, to avoid the boundary errors.

### 3.2 Image modelling and MAP estimation

The statistical model of a noisy image can be written as

\[
z(k, l) = x(k, l) + \eta(k, l)
\]

(3)

Here \( z(k, l) \) is the \((k, l)^{th}\) pixel of the noisy image, \( x(k, l) \) the corresponding noise-free image and \( \eta(k, l) \) is independent and identically distributed (i.i.d.) white Gaussian noise with zero mean and variance \( \sigma^2 \). The aim of denoising is to estimate \( k(k, l) \) from \( z(k, l) \).

Let us denote the \( J \)-level DT with generator matrix \( M_A \) and anisotropic ratio of \( n_2: n_1 \) as \( \text{DT}(M_A, n_1, n_2) \). As explained in Section 2, this DT will generate an approximation subband and \( (2^{n_1/n_2} - 1) \) detail subbands. The approximation subband contains the low frequency portion of the image and thus possesses most of the information of the image.
The detail subbands provide the directional information in the image specific to the spatial location resulting in an efficient energy compaction or sparse representation.

Since 2-D DT is a linear operation like 2-D WT, the equation in (2) can be written as

\[ y_{ij}^{(j)}(k,l) = d_{ij}^{(j)}(k,l) + n_{ij}^{(j)}(k,l) \]  

Here \( y_{ij}^{(j)}(k,l) \), \( d_{ij}^{(j)}(k,l) \) and \( n_{ij}^{(j)}(k,l) \) denote the DT at level \( j \) for segment \( i \) of the segmented noisy image, noise free image and noise components respectively. From now onwards, we will omit the superscript \( i \) and index \((k, l)\) for simplicity.

As in the case of 2-D DWT, the DT of an image yields fairly well decorrelated wavelet coefficients. However, these coefficients are not independent. The large-magnitude coefficients tend to occur near each other within subbands, and also at the same relative spatial locations in subbands at adjacent scales and orientations. The noise level decreases rapidly across scales while signal structures are strengthened. To take advantage of this property we model the coefficients using this dependency.

Figure 4  Interscale dependencies of (a) WT coefficients and (b) directionlet transform coefficients along 90° and 45° (see online version for colours)

In the multi scale DT decomposition, each coefficient, except the ones at the finest scale, has a certain number of children. The number of children that each coefficient can have varies depending on the anisotropic ratio of the transform. WT will always have four children, while DT with two horizontal transforms and one vertical transform will have eight children at the next scale. The interscale dependency is shown in Figure 4(a) in case of WT and in Figure 4(b) in case of DT. Here the case of DT(\( M_{\Lambda J}, 2, 1 \)) is considered, where the generator matrix \( M_{\Lambda J} = [d_1 \ d_2]^T \) with \( d_1 = [1 \ 1] \) and \( d_2 = [0 \ 1] \). The dependent children are identified as in the case of standard zero tree structure. Here the children correspond to the same spatial location and orientation as the parent coefficient. In the case of WT, the set of children is isotropic and aligned along the horizontal and vertical directions. On the other hand, in the case of DT, since directionlets are anisotropic and oriented in different directions, the corresponding children are grouped in anisotropic and oriented sets. The corresponding parent-child-relation is illustrated in Figure 5. The anisotropy and orientation of the sets of children is the same as that used in the
construction of directionlets. But it still retains the property of grouping coefficients across scales that belong to the same spatial location.

**Figure 5** The parent child relationships for DT($M_{\Lambda J}, 2, 1$) (see online version for colours)

Notes: The subbands at the coarser scale $j + 1$ are defined across the lattice $\Lambda_{j+1}$ with the generator matrix $M_{\Lambda_{j+1}}$. Each coefficient has a set of children at the next finer scale $j$.

Let us consider the coefficients of undecimated DT, DT($M_{\Lambda J}, n_1, n_2$) at scale $j$. Since there is one to one dependency between the adjacent scale coefficients, each DT coefficient at scale $Z$ will have one parent at scale $(j + 1)$. If we consider this interscale dependency in the model in equation (4), it can be written as

$$y_j = d_j + n_j$$

where $y_j = (y_{j+1}, y_j)$, $d_j = (d_{j+1}, d_j)$, and $n_j = (n_{j+1}, n_j)$, are the interscale dependence vectors of noisy, noise-free and noise coefficients respectively.

Our aim is to estimate $d_j$ from the noisy observation $y_j$. This can be done by Bayesian MAP estimator. However it may be noted that in the case of decimated DT, since the coefficients in a subband are decorrelated, the (i.i.d) models are largely justified and thus the derivation of a MAP estimator is facilitated. However undecimated DT decomposition generates redundant representation, and there are correlations between the decomposition coefficients. For example, at the first level of decomposition, the odd and even coefficients in each direction are correlated. So in this case the realistic statistical modelling of coefficients is far more difficult. To solve this issue as explained by Chang et al. (2000), we can separate the coefficients into four sets of uncorrelated coefficients, namely $y'_j(2k, 2l)$, $y'_j(2k + 1, 2l)$, $y'_j(2k, 2l + 1)$ and $y'_j(2k + 1, 2l + 1)$. For the $s^{th}$ level decomposition, the coefficients can be separated into $(2^{s+2})^4$ sets, each containing uncorrelated coefficients, and they are i.i.d within each set as well. This approach lets us still use the independent noise assumption and circumvent the issue of denoising correlated signal coefficients with correlated noise.
Image denoising based on adaptive spatial segmentation

The typical MAP estimate of $d_j$ is

$$\hat{d}_j (y_j, y_{j+1}) = \arg \max R_{d_j} (d_j | y_j)$$

(6)

$$\hat{d}_j (y_j, y_{j+1}) = \arg \max \left[ P_{d_j} (d_j | y_j) : p_d (d_j) \right]$$

(7)

$$\hat{d}_j (y_j, y_{j+1}) = \arg \max \left[ p_n (y_j - d_j) : p_d (d_j) \right]$$

(8)

Sendur and Selesnick (2002a) have developed a bivariate shrinkage function based on the parent child relationship in the wavelet domain. The same scheme can be extended to the directionlet domain as DT is an extended WT operation. Here we have to take the joint probability density functions for the noise and signal coefficients by taking into account the parent child relationship. Since the noise is assumed to be independent and identically distributed the joint probability function of noise vector $n_j = (n_j, n_{j+1})$ can be written as a bivariate pdf as

$$p_n (n_j) = \frac{1}{\left( 2\pi \sigma_n \right)^{3/2}} e^{-\frac{\left( n_j^2 + n_{j+1}^2 \right)}{2\sigma_n^2}}$$

(9)

It is well known that WT coefficients of natural images have highly non-Gaussian statistics. Let $d_{j+1}$ is the directionlet coefficient at the same position as $d_j$, but at the next coarser scale, then $d_{j+1}$ can be considered as the parent of $d_j$. The joint probability density function for the signal coefficient vector $d_j = (d_j, d_{j+1})$ can be defined as

$$p_d (d_j) = \frac{3}{\left( 2\pi \sigma \right)^{4/2}} e^{-\frac{\left( d_j^2 + d_{j+1}^2 \right)}{4\sigma^2}}$$

(10)

The model fitting performance of this proposed model to the statistics of natural images is analysed here. The histogram of Lena and Barbara images at two different subbands and levels are considered here and are shown in blue colours in Figures 6 and 7 respectively. The proposed estimated pdf are also shown in red colour in each figure. These figures show that the proposed model provides a very good fit to the histogram of directionlet transform coefficients at different subbands and levels of these images. Also, for a quantitative evaluation of the closeness of fit, the Kullback-Leibler (KL) distance proposed by Uso et al. (2007) is used to measure the difference between normalised histogram and the estimated pdf. It is a non-symmetric measure of the difference between two probability distributions $P$ and $Q$. Specifically, the KL divergence of $Q$ from $P$, denoted as $D_{KL}(P || Q)$, is a measure of the information lost when $Q$ is used to approximate $P$. For discrete probability distributions $P$ and $Q$, the KL divergence of $Q$ from $P$ is defined to be

$$D_{KL} (P || Q) = \sum_i \ln \left( \frac{P(i)}{Q(i)} \right) P(i)$$

(11)
The KL distance of the proposed pdf in (10) and the histograms of DT coefficients at different subbands and resolution levels of Lena and Barbara images are given in Table 1. This also shows that the proposed model has very good model fitting performance. Based on this analysis, the proposed model can very well be used in the MAP estimation process.

Using equation (8) with (9) and (10), the MAP estimator of $d_j$ is derived as

$$
\hat{d}_j(y_j, y_{j+1}) = \frac{\sqrt{y_j^2 + y_{j+1}^2} - \frac{5\sigma_n^2}{\sigma}}{\sqrt{y_j^2 + y_{j+1}^2}} \cdot y_j
$$

(12)

<table>
<thead>
<tr>
<th>Image</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 1</th>
<th>Level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>3.54</td>
<td>3.19</td>
<td>2.98</td>
<td>3.08</td>
</tr>
<tr>
<td>Barbara</td>
<td>2.36</td>
<td>3.16</td>
<td>2.09</td>
<td>2.96</td>
</tr>
</tbody>
</table>

Table 1  Values of the K-L distance between the normalised histogram and estimated pdf

Figure 6  Histogram of the DT coefficients of Lena image and the estimated bivariate pdf
(a) HHH subband at level 1 (b) HHH subband at level 2 (c) HLH subband at level 1 (d) HLH subband at level 2 (see online version for colours)
Here $y_j$ and $y_{j+1}$ are the noisy DT coefficients at two adjacent scales, $\sigma_n^2$ is the noise variance and $\sigma_s^2$ is the marginal variance. So if we know the noise and signal variances, the estimate of the noise free DT coefficients can be computed from the noisy observations of the same coefficients at adjacent scales.

### 3.3 Signal variance estimation through context modelling

The signal variance is estimated here using context modelling technique (Chang et al., 2000). The coefficients in a subband are first grouped in to four groups as explained in Section 3.2 to avoid problem due to the redundant set of coefficients generated by the decomposition with undecimated DT. The coefficients from each of these groups are further classified based on context modelling to differentiate and gather coefficients with some similarities, but not necessarily spatially adjacent. Consider one particular group from a subband with $N$ number of coefficients $y_j[k, l]$. Here the context value, $c_j(k, l)$ of noisy DT coefficient, $y_j[k, l]$ is calculated as the weighted average of the absolute values of its neighbours.

$$c_j(k, l) = \left( \sum_{j'} w_j^{k,j} \right)$$  \hspace{1cm} (13)
where $u_{j,l}^j$ is a $1 \times 8$ vector whose elements are the absolute value of $y_{j[k,l]}$'s eight nearest neighbours. Here the parent coefficient at scale $j + 1$ is not considered as it is already accounted in the bivariate shrinkage function. Here $w_j$ is a $1 \times 8$ weight vector, which can be found by using the least squares estimation as

$$w_j = \left( U_j^T U_j \right)^{-1} U_j^T y_j$$

where $U_j$ is a $N \times 8$ matrix with each row being $u_{j,l}^j$ and $y_j$ is an $N \times 1$ vector containing all elements of $y_{j[k,l]}$. Now the context values are sorted in ascending order and the corresponding DT coefficients are clustered. The DT coefficients with similar natures are now arranged together and are assumed to have same statistics. Now the variance of a DT coefficient $y_{j[k_0,l_0]}$ can be estimated from $y_{j[k,l]}$ whose context $c_{j[k,l]}$ falls near to it. The pixel values can be selected by $L$ closest points in value above $c_{j[k_0,l_0]}$ and $L$ closest points in value below, resulting in a total of $2L$ points. These $2L$ points can be considered as a cluster $\Theta_j^l$. The variance of DT coefficient can be estimated from this cluster as

$$\hat{\sigma}_j^2 = \text{Max} \left( \frac{1}{2L} \sum_{y_{j[k,l]} \in \Theta_j^l} y_j^2 - \sigma_n^2, 0 \right)$$

where $\sigma_n^2$ is the noise variance which needs to subtract because $y_{j[k,l]}$ are the noisy observations and the noise is independent of the signal. The noise variance can be computed from the high frequency sub band coefficients at the finest scale in the decomposition (HHH1) as:

$$\sigma_n = \frac{\text{Median}(\{y_i\})}{0.6745}, \text{ where } y_i \in \text{HHH}_1$$

The selection of the value of $L$ is very important here. Too small or too large values adversely affect the performance. In the proposed scheme we have chosen this value as 20.

### 3.4 Proposed denoising algorithm

The full denoising algorithm proposed in the paper can be summarised in three steps as follows:

**Step 1** Adapative spatial segmentation and identification of dominant direction

- The input image is first divided into four equally sized segments.
- Compute DirVar for each spatial segment along eight directions using equation (2).
- If the DirVar for all directions lies in within 5%, then leave the segment as such, else check whether there are more than two dominant directions in that segment. If it is so, then further divide it into four equally sized segments and go back to the previous step until a pre defined segmentation depth is reached.
Step 2  Multi scale directionlet transform computation and estimation of noise free coefficients

- apply multi scale undecimated DT to the segments along the optimal pair of directions
- for segments with no apparent dominant direction, compute the DT along (0°, 90°)
- for each DT coefficient, compute the marginal variance $\sigma^2$ using equation (15)
- estimate the new noise free DT coefficient using equation (12).

Step 3  Reconstruction

- Reconstruct the image from the modified DT coefficients using directional information and inverse DT.

4  Experimental results

The performance of the proposed algorithm was tested by using standard eight-bit grey scale images Lena, Boat and Barbara of size 512 × 512. The Lena and Boat images contain smooth regions and edges while the Barbara image has abundant inhomogeneous structures. These images were contaminated with zero mean white Gaussian noise with variance 10, 15, 20, 25 and 30. The performance of the proposed method was compared with the published results of denoising schemes available in the literature such as Bayes shrink (Chang et al., 2000), bivariate shrinkage in DWT and DTCWT domain (Sendur and Selesnick, 2002b) and DTCWT with three scales of dependency (Chen et al., 2012).

In the proposed method the image was quad tree segmented based on the content directionality and a maximum segmentation depth of 32 × 32 was fixed. The segmentation depth greatly affects the quality of processed images. If it is too small, the algorithm may not work well and may lead to unwanted block-like artifacts. On the other hand, if the segment size is too large, the region containing image features also become large with relatively more number of directions. This will adversely affect the denoising performance. After segmentation, each segment was decomposed using undecimated DT up to four levels with Daubechies’ wavelet with eight vanishing moments.

The performance improvement of the proposed scheme was quantified in terms of peak signal to noise ratio (PSNR), which was computed using the following formula:

$$
PSNR = 10 \log_{10} \left( \frac{N \times 255^2}{\sum_{k,l} (I_{den}(k,l) - I_{org}(k,l))^2} \right)
$$

(17)

where $I_{den}$ is the denoised image, $I_{org}$ the original image (before adding noise) and $N$ is the number of pixels in the image. The values obtained are given in Table 2 for different noise realisations. The PSNR values provided are the average of ten different noise realisations for each standard deviation.
Table 2  PSNR (dB) values of Lena, Boa and Barbara images with different denoising algorithms

<table>
<thead>
<tr>
<th>Variance</th>
<th>Noisy</th>
<th>Bayes shrink</th>
<th>DTCWT-3</th>
<th>Bivariate with WT</th>
<th>Bivariate with DTCWT</th>
<th>Proposed scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>28.13</td>
<td>33.32</td>
<td>35.32</td>
<td>34.37</td>
<td>35.32</td>
<td>35.70</td>
</tr>
<tr>
<td>15</td>
<td>24.61</td>
<td>31.41</td>
<td>33.60</td>
<td>32.50</td>
<td>33.64</td>
<td>33.96</td>
</tr>
<tr>
<td>20</td>
<td>22.11</td>
<td>30.17</td>
<td>32.36</td>
<td>31.20</td>
<td>32.41</td>
<td>32.72</td>
</tr>
<tr>
<td>25</td>
<td>20.17</td>
<td>29.22</td>
<td>31.38</td>
<td>30.15</td>
<td>31.38</td>
<td>31.74</td>
</tr>
<tr>
<td>30</td>
<td>18.59</td>
<td>28.48</td>
<td>30.56</td>
<td>29.41</td>
<td>30.51</td>
<td>30.90</td>
</tr>
<tr>
<td>10</td>
<td>28.13</td>
<td>31.80</td>
<td>33.23</td>
<td>32.41</td>
<td>33.12</td>
<td>33.71</td>
</tr>
<tr>
<td>15</td>
<td>24.61</td>
<td>29.87</td>
<td>31.35</td>
<td>30.54</td>
<td>31.34</td>
<td>31.79</td>
</tr>
<tr>
<td>20</td>
<td>22.11</td>
<td>28.48</td>
<td>30.01</td>
<td>29.20</td>
<td>30.09</td>
<td>30.37</td>
</tr>
<tr>
<td>25</td>
<td>20.17</td>
<td>27.40</td>
<td>28.98</td>
<td>28.13</td>
<td>29.02</td>
<td>29.45</td>
</tr>
<tr>
<td>30</td>
<td>18.59</td>
<td>26.60</td>
<td>28.16</td>
<td>27.30</td>
<td>28.33</td>
<td>28.50</td>
</tr>
<tr>
<td>10</td>
<td>28.13</td>
<td>30.86</td>
<td>33.66</td>
<td>32.23</td>
<td>33.35</td>
<td>34.00</td>
</tr>
<tr>
<td>15</td>
<td>24.61</td>
<td>28.51</td>
<td>31.49</td>
<td>30.00</td>
<td>31.28</td>
<td>31.87</td>
</tr>
<tr>
<td>20</td>
<td>22.11</td>
<td>27.13</td>
<td>29.97</td>
<td>28.35</td>
<td>29.81</td>
<td>30.41</td>
</tr>
<tr>
<td>25</td>
<td>20.17</td>
<td>26.01</td>
<td>28.78</td>
<td>27.17</td>
<td>28.58</td>
<td>29.20</td>
</tr>
<tr>
<td>30</td>
<td>18.59</td>
<td>25.16</td>
<td>27.84</td>
<td>26.30</td>
<td>27.62</td>
<td>28.12</td>
</tr>
</tbody>
</table>

From the results it is clear that the proposed method outperformed the compared schemes for all the test images. It outperforms the classical BayesShrink by more than 2 dB on Lena and Boat images and more than 3 dB on Barbara images. The visual quality of the denoised images is also used to evaluate the performance. The difference in visual quality is mainly in the preservation of image edges and textures. In Figures 8 and 9, extracted portions of the Lena and Barbara images are compared with bivariate shrinkage in DTCWT domain. The visual quality of the denoised image using the proposed algorithm is evidently better because of sharper edges and texture. To emphasise this gain, portions of the denoised images of Barbara containing the stripes of the scarf and Lena containing the lines of the hat are zoomed in for both the methods. It can be noticed that the stripes in the scarf of Barbara image and the cross lines in the hat of Lena image are more clear and distinct in the denoised images of the proposed scheme. In Figure 10, two different regions of Barbara image containing smooth and edge areas are extracted and the performance is compared with bivariate shrinkage in DTCWT domain. The interscale dependency modelling in directionlet domain for high frequency images like Barbara gives better results as it adaptively captures the oriented features like edges and contours in this image which correspond to high frequency information.
Figure 8  Comparing the results of DTCWT and directionlet-based denoising methods, for part of Lena image, (a) noise free image (b) noisy image with $\sigma_n = 20$ (c) denoised image using bivariate shrinkage in DTCWT (d) denoised image using the proposed scheme (e) selected enlarged portion of ‘(c)’ (f) selected enlarged portion of ‘(d)’
Figure 9  Comparing the results of DTCWT and directionlet-based denoising methods, for part of Barbara image, (a) noise free image (b) noisy image with $\sigma_n = 20$ denoised image using bivariate shrinkage in DTCWT (d) denoised image using the proposed scheme (e) selected enlarged portion of (c) (f) selected enlarged portion of (d)
Figure 10  Image denoising results of two regions of Barbara image

Note: The images are arranged from left to right in order from noise free image, noisy image with $\sigma_n = 20$ denoised image using DTCWT and denoised image using proposed scheme.

A comparison of the computational efficiency of the proposed algorithm with the bivariate shrinkage in DTCWT domain is carried out here. The computational cost of both these algorithms consists of the cost of image decomposition using the respective transform, filtering process for the estimation of noise free coefficients and the reconstruction. In Bivariate shrinkage in DTCWT domain there are four numbers of 1-D filtering in each level. In the proposed scheme there is only two 1-D filtering in each stage as it uses isotropic transform along the dominant directions. The same number of filtering is required in the reconstruction phase also. The extra computation is in the spatial segmentation and identification of dominant directions. This adds to the computational complexity of DT-based applications. So there is a tradeoff between the computational efficiency and denoising performance in DT-based schemes.

5 Conclusions

In this paper an effective image denoising technique based on bivariate shrinkage in directionlet domain is presented. The directionlets are constructed adaptively so that the chosen directions are maximally aligned with locally dominant directions across image. Because of the alignment, the transform generates a sparser representation with reduced energy in the high-pass subbands allowing for a more robust estimation of the noise free coefficients. Here an efficient model is developed which uses the joint statistics of the DT coefficients of natural images and a nonlinear threshold function is derived using Bayesian MAP estimator. The signal variance computation based on context modelling of the DT coefficients causes performance improvement by way of incorporation of intra scale dependency in the estimation. The experimental results with the commonly used test images show that the new algorithm achieves competitive performance with the existing algorithms based on DWT and DTCWT. The denoised images achieve better visual quality by preserving the edges and textures efficiently. Adaptive spatial
segmentation and simple 2-D separable transform-domain filtering reduces the computational complexity of the algorithm. A more complex spatial segmentation, which is capable of extracting the dominant directions in images with more flexibility, may further improve the performance of the proposed scheme. The identification of dominant directions in the case of images contaminated with large amount of noise is a challenging task. The estimation of dominant directions based on the computation of directional variance may not yield a good result in such cases. The future work is aimed for addressing these aspects.

References


Olshausen, B.A. and Field, D.J. (1996) ‘Emergence of simple cell receptive filed properties by 

pp.117–125.

exploting interscale dependency’, *IEEE Trans. Signal Process.*, Vol. 50, No. 11, 
pp.2744–2756.


Sethunadh, R. and Thomas, T. (2013) ‘SAR image despeckling in directionlet domain based on 


No. 12, pp.4158–4171.

pp.1761–1773.
