13

Suppression of Correlated Noise

Jan Aelterman, Bart Goossens, Aleksandra Pizurica and Wilfried Philips
Ghent University, TELIN-IPI-IBBT
Belgium

1. Introduction

Many signal processing applications involve noise suppression (colloquially known as denoising). In this chapter we will focus on image denoising. There is a substantial amount of literature on this topic. We will start by a short overview:
Many algorithms denoise data by using some transformation on the data, thereby considering the signal (the image) as a linear combination of a number of atoms. For denoising purposes, it is beneficial to use such transformations, where the noise-free image can be accurately represented by only a limited number of these atoms. This property is sometimes referred to as sparsity. The aim in denoising is to detect which of these atoms represent significant signal energy from the large amount of possible atoms representing noise.
A lot of research has been performed to find representations that are as sparse as possible for 'natural' images. Examples of such representations are the Fourier basis, the Discrete Wavelet Transform (DWT) (Donoho, 1995), the Curvelet Transform (Starck, 2002), the Shearlet transform (Easley, 2006), the dual-tree complex wavelet transform (Kingsbury, 2001; Selesnick, 2005), … Many denoising techniques designed for one such representation can be used in others, because the principles (exploiting sparsity) are the same. Without exception, these denoising methods try to preserve the small amount of significant transform coefficients, i.e the ones carrying the information, while suppressing the large amount of transform coefficients that only represent noise. The sparsity property of natural images (in its proper transform domain) ensures that there will be only a very small amount of significant transform coefficients, which allows to suppress a large amount of the noise energy in the insignificant transform coefficients. Multiresolution denoising techniques range from rudimentary approaches such as hard or soft thresholding of coefficients (Donoho, 1995) to more advanced approaches that try to capture some statistical significance behind atom coefficients by imposing appropriate prior models (Malfait, 1997; Romberg, 2001; Portilla, 2003; Pizurica, 2006; Guerrero-Colon, 2008; Goossens, 2009;).
Another class of algorithms try to exploit image (self-) similarity. It has been noted that many images have repetitive features on the level of pixel blocks. This was exploited in recent literature through the use of statistical averaging schemes of similar blocks (Buades 2005; Buades, 2008; Goossens, 2008) or grouping of similar blocks and 3d transform domain denoising (Dabov, 2007).

In practice, processes that corrupt data can often not be described using a simple additive white gaussian noise (AWGN) model. Many of these processes can be modelled as linear filtering process of a white Gaussian noise source, which results into correlated noise. Some correlated noise generating processes are described in section 2. The majority of the mentioned denoising techniques are only designed for white noise and relatively few techniques have been reported that are capable of suppressing correlated noise. In this chapter, we present some techniques for noise estimation in section 4 and image modelling in section 3, which form the theoretical basis for the (correlated) noise removal techniques explained in section 5. Section 6 contains demonstration denoising experiments, using the explained denoising algorithms, and presents a conclusion.

2. Sources of Correlated Noise

2.1 From white noise to correlated noise

In this section, the aim is to find a proper description of correlated noise. Once established, we will use it to describe several correlated noise processes in the remainder of this section. Since the spatial correlation is of interest, rather than time/spatially-varying noise statistics, we will assume stationarity throughout this chapter. Stationarity means that the autocorrelation function only depends on the relative displacement between two pixels, rather than their absolute positions. This is evident from (1), a random process generating samples f(n) is called white if it has zero mean and a delta function as autocorrelation function $r_f(n)$:

$$\begin{aligned} &E\left[f(n)\right]=0 \\ &r_f(n)=E[f(m)\overline{f(n+m)}]=\partial(n) \end{aligned} \tag{1}$$

The Wiener–Khinchin theorem states that the power spectral density (PSD) of a (wide-sense stationary) random signal f(n) is the Fourier transform of the corresponding autocorrelation function r(n):

$$R(\omega)=\sum_{n=-\infty}^{\infty} r(n)e^{-j\omega n} \tag{2}$$

This means that for white noise, the PSD is equal to a constant value, hence the name white (white light has a flat spectrum). When a linear filter h(n), with Discrete Time Fourier Transform (DTFT) H(ω) is applied (often inadvertently) to the white noise random signal, the resulting effect on the autocorrelation function and PSD of $f'(n)=f(n)*h(n)$ is:

$$\begin{aligned} &r'(n)=E[f'(m)\overline{f'(n+m)}]=\sum_{m=-\infty}^{\infty} h(m)\overline{h(n+m)} \\ &R'(\omega)=H(\omega)\overline{H(\omega)}=\left|H(\omega)\right|^2 \end{aligned} \tag{3}$$

This result shows that the correlated noise PSD R′(ω) is the squared magnitude response of the linear filter DTFT, hence one can think of correlated noise as white noise subjected to linear filtering. In analogy with the term 'white noise' this is sometimes referred to as 'colored noise'. In the following sections, some real world technologies will be explained from the perspective of noise correlation.

2.2 Phase Alternating Line (PAL) Television

PAL is a transmission standard used in colour analogue broadcast television systems. Dating back to the 1950s, there are several bandwidth saving techniques that are very nice in their own right, but are responsible for the noise in PAL television. One is the deinterlacing mechanism (Kwon, 2003). Another is the use of a different modulation and filtering schemes. We will restrict us here to show the PSD of a patch of noise from a PAL signal broadcast:

Fig. 1. noisy PAL broadcast of a sports event and PSD of noise in a green color channel of the PAL broadcast.

It is clear that the noise here is almost cut off horizontally, leading to stripe like artifacts and there is significant energy in the lower vertical frequencies, leading to vertical streaks. It is therefore naive to assume noise in PAL/NTSC television to be white.

2.3 Demosaicing

Modern digital cameras use a rectangular arrangement of photosensitive elements. This matrix arrangement allows the interleaving of photosensitive elements of different color sensitivity. This interleaving allows sampling of full color images without the use of three matrices of photosensitive elements. One very popular arrangement is the Bayer pattern (Bayer, 1976), shown in figure 2.

Fig. 2. Bayer mosaic pattern of photosensitive elements in a camera sensor

There exist a wide range of techniques for reconstructing the full color image from mosaiced image data. A thorough study of these techniques is beyond the scope of this chapter. Instead, we compare the simplest approach with one state of the art technique, from the viewpoint of noise correlation.

Since all of these techniques perform interpolation in some way, we are confident that the conclusion will be similar for all demosaicing techniques. When using patches of white noise as input data for an image mosaic sensor, the PSD of the demosaicing noise is shown in figure (3)-(4). Because white noise is used as input for a demosaicing system, the reconstruction will be a color image of correlated noise. Since we are assuming additive noise, this will be a good model for type of noise that is encountered in color digital camera images. In figure (3), we compared the simplest demosaicing approach, bilinear demosaicing, for both the green and the red color plane. Since the red data is subsampled more, the demosaicing treats both color planes differently, explaining the difference in PSD bandwidth. Figure (4) shows the result using a state-of-the-art demosaicing technique (Aelterman, 2009). There is no visible difference between the green and blue plane PSD here, because the algorithm works in a luminance/chrominance approximation rather than red/green/blue space.

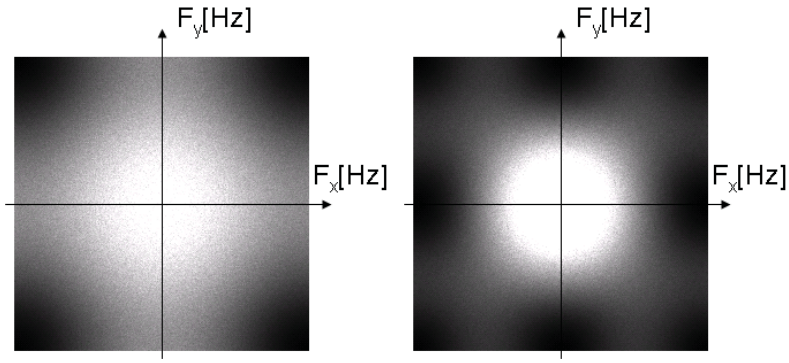


Fig. 3. PSD of the green channel (left) and a red/blue channel(right) from the demosaiced white noise patch using bilinear demosaicing.

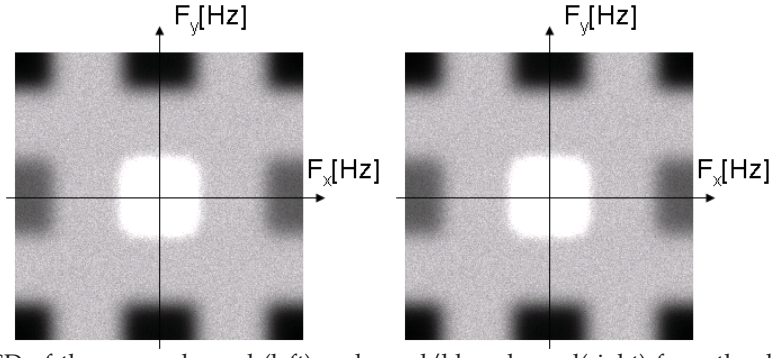


Fig. 4. PSD of the green channel (left) and a red/blue channel(right) from the demosaiced white noise patch using the technique described in (Aelterman, 2009)

The demosaicing experiments in figure 4 and 5 also show that the low pass part of the PSD is brighter, meaning it represents more noise energy. This is explained through the tendency

of demosaicing algorithms to favour smooth color patches, as this is an accurate model for natural images.

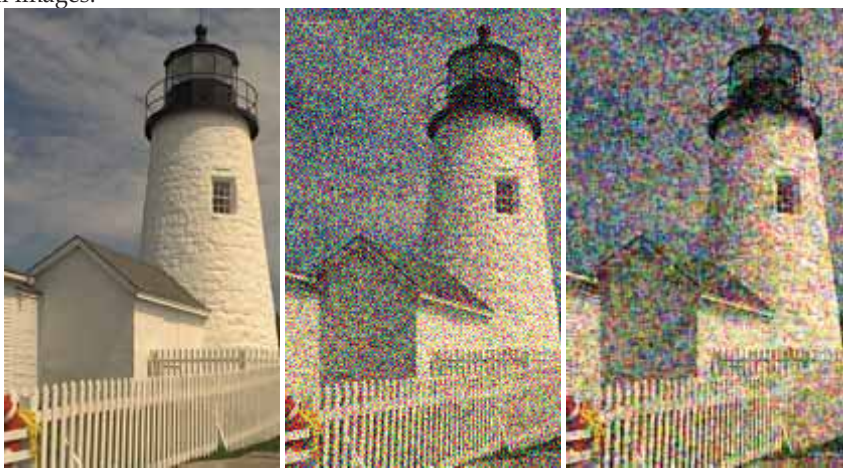


Fig. 5. The full color lighthouse image (left) corrupted with white noise (middle) and with correlated noise due to bilinear demosaicing (right)

2.3 Thermal Camera's

Many thermal cameras are based on the push broom or whisk broom principle. Only very few infrared sensors are used, which have to be reused through an optics system in order to scan the different pixels positions, creating the complete image. This raster scan principle is illustrated in figure 6.

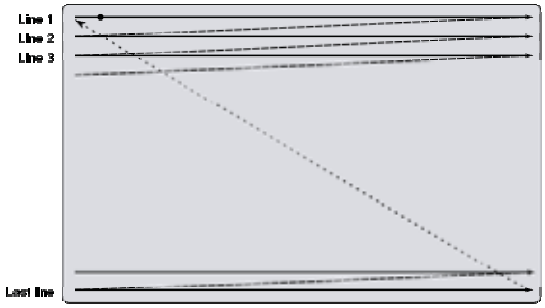


Fig. 6. Example of a raster scanned image pattern (whisk broom imaging)

The downside of such imaging principles is that they sometimes exhibit streaking noise artifacts, which can be attributed to the sensor and sampling circuitry. Since pixel intensities at different spatial positions are acquired using the same sensor, temporal correlation in the noisy sensor data results in spatial noise correlation in the acquired image. For thermal cameras, noise can be approximated using a $1/f$ frequency dependency of the noise (Borel, 1996). This type of 'pink' noise is very common in electronic devices and becomes apparent when reusing image sensors for different pixels in the image at high sample rates.

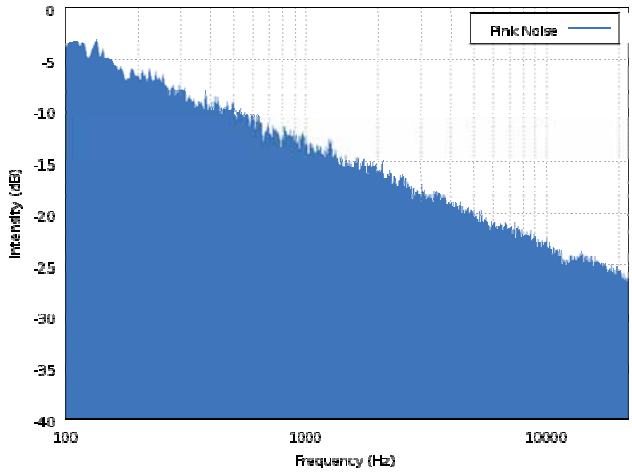


Fig. 7. PSD of 'pink noise', an approximate model of noise in some types of electronic devices through time

Pink noise can easily be simulated, by filtering a time sequence of white pseudorandom numbers with a 1/f filter characteristic and then adding those noise samples to the signal values in a raster scan pattern. Figure 8 shows this type of noise on the Einstein image.

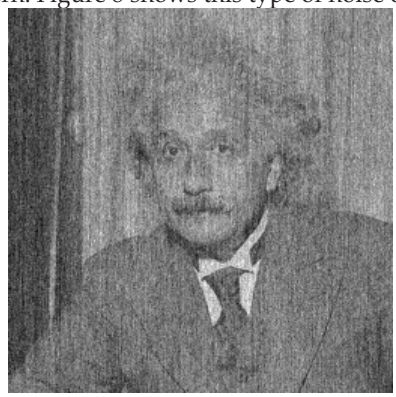


Fig. 8. The Einstein image with simulated whisk broom imaging 1/f noise.

The streaking noise artifacts are clearly visible.

2.4 Magnetic Resonance Imaging

Magnetic Resonance Imaging is a non-invasive imaging technique. The Signal to Noise ratio (SNR) is heavily dependent on scan time. Longer scans are not only less comfortable for the patient, it is also less economical for the scanner operator and it increases the chance of involuntary patient movement. Because MRI technology acquires a Fourier transformed version of the image, rather than the image itself, patient movement during a scan translates

to 'echo'-like artifacts in the image, which is much more detrimental to the diagnostic image quality than simple blurring in conventional photography.
These facts are a major motivation for scanner manufacturers, who have created a wide range of tricks and technologies that allow an MRI scan to be made faster or less susceptible to motion artifacts. Noise in MRI is traditionally considered white (e.g. Nowak, 1999; Pizurica, 2003), and this is indeed a good noise model for theoretical MRI, but in practice, almost all clinical MRI acquisition devices use one or more acceleration techniques. As an illustration we now explain a few:
In K-space subsampling, bands of the spectrum of the signal are simply not scanned. Elliptical filtering is similar, where only elliptical area around the K-space center is sampled. An illustration and example of this principe is shown in figure 9.

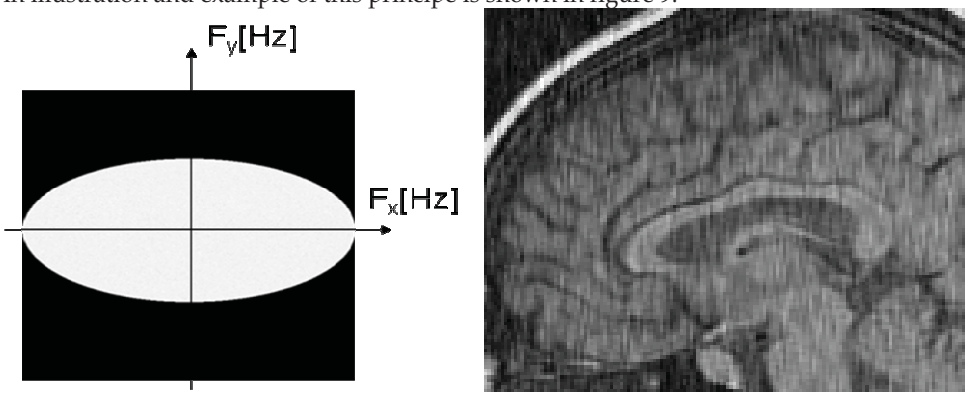


Fig. 9. (left) PSD of K-space subsampled and elliptically filtered noise (right) Brain MRI with simulated K-space subsampled and elliptically filtered MRI noise

It is clearly visible that for this situation, the noise creates stripe-like artifacts rather than being statistically independent from one pixel to another.

3. Describing natural images

3.1 Sparsity

When generating an image using purely random generators, chances are it will look like random noise. To the authors' knowledge, nobody ever succeeded in generating an image this way that looks like a 'natural' picture, which could have come from a camera. This is because what people consider 'natural' images is a very small subset of all possible images. Natural images have specific properties. These properties have been the object of intensive study throughout the years and are of interest here, as it is what will enable the statistical separation of noise from signal in images.
An easily verifiable, intuitive, property of natural images is that they are smoothly varying signals, which give low frequency responses, delineated by lines (edges). This is somewhat quantified by the inverse power scaling law, natural image PSD's are inversely proportional in magnitude to the squared modulus of the spatial frequency variable (Ruderman, 1994; Field, 1987):

$$R'(\omega) \propto \frac{1}{\omega^{2+\eta}} \tag{4}$$

Where η is a small number (|η|<1). This observation, coupled with the mathematical elegance of the Gaussian distribution, motivated early image processing engineers to model both image and noise as Gaussian distributed, in some transform domain. The Minimization of the Mean Squared Error Bayesian risk estimator (MMSE) denoising solution is the well known Wiener filter. However, the Gaussian distribution does not account for the relatively large number of outliers when considering e.g. wavelet filter responses. In fact, as a consequence of thinking of images as smooth regions delineated by edges, most coefficients will be far smaller than predicted by the Gaussian model and a small number will be really big.

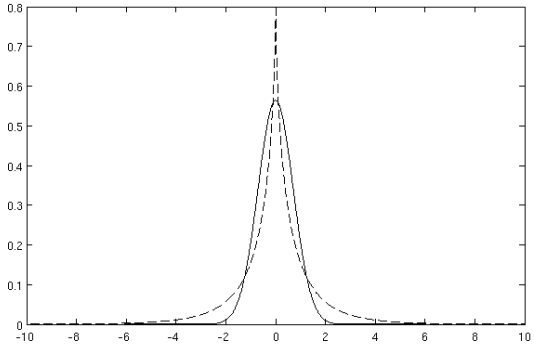


Fig. 10. Comparison of a Gaussian distribution (full line) with a more heavily tailed distribution, which is a better model for the transform coefficient distribution of natural images

This property, sometimes referred to as sparsity, is better modeled by a more heavily tailed distribution model, a comparison between both types of prior distributions is shown in figure 11. The next section gives some examples.

3.2 Marginal statistics of multiresolution image transformation coefficients

It is clearly visible in multiresolution decompositions that marginals distributions of coefficients exhibit highly non Gaussian behaviour. In fact, more leptokurtic distributions account better for the heavy tailed marginal distributions that one encounters in practice. For this reason, other distributions, such as the Laplacian (also known as the exponential power distribution) or Generalized Laplacian (Mallat, 1989; Pizurica, 2006) have been applied in more effective techniques. The Generalized Laplacian distribution is described in (5), with λ and υ called the scale and shape parameter and Γ() the gamma function.

$$p(\beta) = \frac{\lambda\upsilon}{2\Gamma\left(\frac{1}{\upsilon}\right)} \exp\left(-\left|\lambda\beta\right|^{\upsilon}\right) \tag{5}$$

The Gaussian Scale Mixtures (GSM) (Wainwright, 2000) is a more elegant alternative, because it can be extended more easily to the multivariate situation, which is advantageous when considering spatial noise correlation. A random vector **β** is a GSM when it can be written as:

$$\beta \overset{d}{=} \sqrt{z}U \tag{6}$$

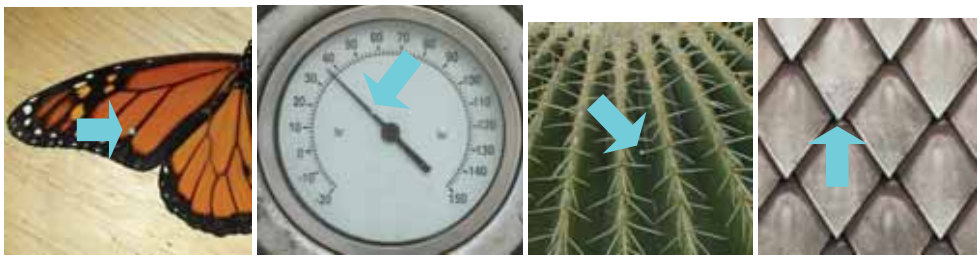
Where **U** is a Normally distributed random vector with mean **0** and covariance matrix **C**, z>0 is a random scalar variable, independent of **U** and superscript d means 'equal in distribution'. The marginal distribution of **β** can then be written as:

$$p(\beta) = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{N/2}\left|z\ C\right|^{1/2}} \phi(z) \exp\left(-\frac{\beta^t C^{-1}\beta}{2z}\right) dz \tag{7}$$

Both the Generalized Laplacian and the GSM model are parameterized, which means that some (often empirical) parameter estimation is required for the parameters. For the Generalized Laplacian, this can be done by fitting the sample kurtosis and variance, while for the GSM covariance matrix and hidden multiplier z distribution, there exist an expectation maximization (EM) solution (Portilla, 2003). It is also possible to assume a fixed distribution for the hidden multiplier z, which will enhance the kurtosis of the marginal distribution p(β) in order to have a more heavily tailed distribution that corresponds more to the model from section 3.1 e.g. Jeffey's prior (p(z)=1/z).

3.2 Selfsimilarity

Another property of natural images is that they often contain a lot of redundant information. Natural images contain a lot of repeating structures and texture. Often, similarity can also be encountered at different scales. Some images of plants or art can be surprisingly well described using fractals, meaning that they can be described as built up of downscaled copies of themselves. This suggests that there is a lot of similarity in an image, a property called selfsimilarity. Since noise is typically more spatially independent, it is a good idea to find a description for self similarity, which can then be exploited in a denoising algorithm. This is done by finding similar image patches for every pixel in the image.



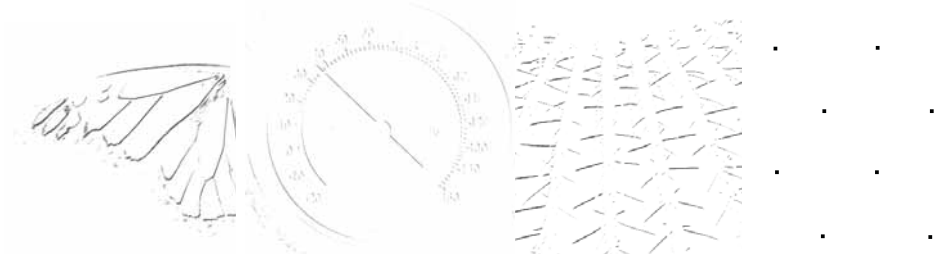


Fig. 11. Several natural images (top), with a similarity map (MSE) for the cyan pixel (bottom). Note how there is a large number of similar pixels for natural images.

Finding similar patches in an image can be done by block matching: For some neighborhood window size L, the mean squared error is taken between every neighborhood in the image and the neighborhood around the current pixel. Similar pixels exhibit a low MSE and the selfsimilarity property ensures that there will be plenty of such similar patches in a natural image. Figure 11 shows the similarity values for every pixel in the image, with respect to the image indicated by the arrow for a neighborhood size of 3x3. Black means a high similarity (low MSE). Note that for the cactus image, all the needles are found to be similar, even though the perspective causes the needles on the top of the image to be smaller versions of the needles at the bottom, which is an example of where similarity can be found at different scales.

4. Estimating noise Correlation

4.1 Signal-free data

In many applications, noise can be considered additive. Additive noise means that even after linear image transformations, the noise remains additive. In this case, the noise statistics are not signal dependent, and could just as well be estimated from signal-free data. In fact, this is easier, because the estimation will not be hindered by the signal. Many applications allow for the acquisition of signal-free data. This is the case for MRI images, where physical limitations ensure the existence of signal-free regions next to the scanned subject in the MRI image. For an unknown thermal camera type, one could photograph a surface of even temperature. Similar techniques can be used for optical photography and many other imaging modalities. When presented with such signal-free, noisy data, the autocorrelation function can be calculated in a straightforward way.

Since the Discrete Fourier Transform (DFT) can be used as a good approximation for the Discrete Time Fourier Transform (DTFT) in (2) for finite data, it is possible to obtain the autocorrelation function from the power spectral density, which can be estimated from the finite noisy, signal-free input data f(n):

$$F(k) = \sum_{n=0}^{N-1} f(n) e^{-\frac{j2\pi kn}{N}}$$
$$r(n) = \sum_{k=0}^{N-1} F(k)\overline{F(k)} e^{\frac{j2\pi kn}{N}} \tag{8}$$

This way of using the DTFT to estimate the autocorrelation function of a stationary process is computationally more efficient than calculating the autocorrelation function or covariance matrix explicitly through e.g. convolution in the image domain.

4.3 Estimating correlated noise in transform domains

Calculating the pixel (voxel) autocorrelation function / covariance matrix is just part of the solution. Since many denoising algorithms operate in a transform domain, it is necessary to transform the autocorrelation function / covariance matrix as well. A transformed covariance matrix enables the whitening of transform coefficients, which allows the use of a white noise denoising algorithm.

One obvious way to obtain the transformed covariance matrices, is through Monte Carlo simulations. First patches of noise are generated, then transformed into the transform domain and then the noise covariance matrix are calculated for every obtained subband. Doing this for multiresolution transforms with a lot of scales either results in heavy memory requirements or high computational requirements, in order to get meaningful results.

In (Goossens, 2009), a different approach is presented. Many transformations can be viewed as banks of linear filters. Applying a transform filter G(ω) to the signal f'(n) from (3), the PSD R''(ω) becomes:

$$R''(\omega) = G(\omega)R'(\omega)\overline{G(\omega)} \tag{9}$$

When also taking the subsampling step from many shift-variant transformations (such as the discrete wavelet transform) into account, the transform becomes less trivial:

$$R''(2\omega) = \Big(G(\omega)F'(\omega)+G(\omega+\pi)F'(\omega+\pi)\Big)\overline{\Big(G(\omega)F'(\omega)+G(\omega+\pi)F'(\omega+\pi)\Big)} \tag{10}$$

In (10) we see the appearance of crosscorrelation terms $F'(\omega)\overline{F'(\omega+\pi)}$ which seem to indicate that it is not possible to simply filter the autocorrelation function such as in (9). However, considering the stationarity assumption, the following should also hold:

$$R''(2\omega) = \big(G(\omega)F'(\omega) - G(\omega+\pi)F'(\omega+\pi)\big)\overline{\big(G(\omega)F'(\omega) - G(\omega+\pi)F'(\omega+\pi)\big)} \tag{11}$$

The minus sign is caused by a one sample shift prior to subsampling. Now the crosscorrelation terms appear with an inverted sign with respect to (10). Since the stationarity assumption implies equality between (11) en (10) it follows that the crosscorrelation are zero. This means that the calculation of the transform autocorrelation function simplifies to:

$$R''(2\omega) = G(\omega)R'(\omega)\overline{G(\omega)} + G(\omega+\pi)R'(\omega+\pi)\overline{G(\omega+\pi)} \tag{12}$$

Again, the transform autocorrelation functions are simply filtered and subsampled versions of the original autocorrelation function. Noise covariance in subbands of a separable, decimated wavelet transform with an arbitrary number of scales can hence be calculated.

4.4 Full Blind GSM estimation

Many scenarios exist where there is no (sufficient) signal-free data. For this situation, the problem of obtaining accurate noise covariance estimations is not so trivial. The median absolute deviation (MAD) of the image transform coefficients in the finest scale (Donoho, 1994) is a robust estimator for the noise variance for white noise. The reason for using the finest resolution scale is that it contains the smallest signal contribution, which is in turn explained by the tendency of natural images to contain predominantly low frequency content. However, correlated noise is by its very definition frequency dependent, which means an estimate from a high frequency subband cannot be used for other transform subbands. Lower frequency subbands have a larger signal contribution, making noise covariance estimations less accurate.

Since the mentioned image transformations are designed to yield sparse responses for natural images, there are still substantial signal-free regions inside each subband. From these areas, an estimate for the noise covariance could be made in the respective transform subband.

There has been some research to put this principle in practice, through a maximum likelihood formulation of the signal estimation problem (Portilla, 2004). The algorithm relies on the different statistical properties of (correlated) Gaussian noise and natural image signal, which is modelled as a GSM (section 3.2). Considering the signal as a series of neighbourhood vectors of arbitrary size, the vector of noisy coefficients is then written as **y=x+w**. Considering the GSM model, one can also write: $\mathbf{y}=\sqrt{z}\,\mathbf{.U+w.}$ Using conditional distributions on the hidden GSM multiplier z, these observed vectors can be written as:

$$p_y(y) = \int_0^\infty p_{y|z}(y \mid z) p_z(z) dz \tag{13}$$

Now consider, without loss of generality, E[z]=1. This means that the underlying signal covariance $C_x = C_u$. Note that the Gaussianity of the scale mixture components and of the noise means it is very easy to write the conditional probability of the observed vector on the hidden GSM multiplier z:

$$p_{y|z}(y \mid z) = \frac{\exp\left(-y^t \left(zC_u + C_w\right)^{-1} y/2\right)}{(2\pi)^{N/2} \left|zC_u + C_w\right|^{1/2}} \tag{14}$$

From this model, we need to obtain the noise covariance matrix. Note that C_x can be estimated as $\left\lfloor C_y - C_w \right\rfloor_+$ where the '+' sign signifies an operation to only preserve positive eigenvalues since covariance matrices need to be positive semidefinite in order to keep the marginal probability values lower than 1. Consider the case for small values of z in equation (17), then $p_{y|z}(y \mid 0 \le z < dz)dz$ approaches the Gaussian noise distribution. The

conditional $p_{z|y}(0 \le z < dz \mid y)dz$ can then be considered the conditional probability that an observed vector **y** was generated by this noise process. Then, the weighted sample covariance calculation (15) will automatically exclude neighborhoods which are likely to contain signal contributions.

$$C_w^{new} = \frac{\sum_{m=1}^{M} p_{z|y}(0 \le z < dz \mid y_m; C_w^{old}) y_m . y_m^t dz}{\sum_{m=1}^{M} p_{z|y}(0 \le z < dz \mid y_m; C_w^{old}) dz} \tag{15}$$

Note that the calculation of the conditionals on z requires knowledge of C_w. This equation needs to be iterated in an Generalized Expectation Maximization (GEM) (Hastie, 2001) style until convergence to reach the ML estimate for the noise covariance matrix. The weights in (15) are visualized in figure 12 for one subband of a noisy version of the Einstein image.

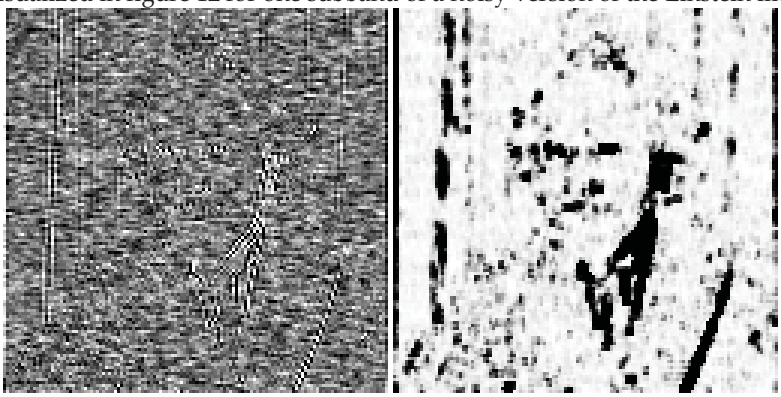


Fig. 12. Noisy subband of the Einstein image (left) weight map for the noise covariance estimation (right) Black corresponds to low weights, note how all the visible image features are detected and excluded from the covariance estimation.

Where the hidden multiplier distribution $p_z(z)$ is unknown, i.e. when a fixed prior distribution, such as Jeffrey's prior or others are not assumed, it is also necessary to estimate $p_z(z)$. Here the choice is to use the classical EM solution for Gaussian scale mixtures:

$$\begin{aligned} p_z^{new}(z) &= \frac{1}{M}\sum_{m=1}^{M} p_{z|y}\left(z \mid y_m; p_z^{old}(z)\right) \\ &= p_z^{old}(z)\frac{1}{M}\sum_{m=1}^{M} \frac{p_{y|z}(y_m \mid z)}{\int_0^{\infty} p_{y|z}(y_m \mid \alpha) p_z^{old}(\alpha) d\alpha} \end{aligned} \tag{16}$$

The combination of (15) and (16) allows to estimate the noise covariance matrices for the subband representations of an image in a transform domain. These can be combined into the

pixel domain covariance matrix or autocorrelation functions using the filter bank technique from section 4.3 and (18) and subsequently be used in correlated noise denoising algorithms.

5. Denoising Algorithms

In this section, some state-of-the-art denoising algorithms are presented. It should be noted that the all of them, except for the NLMS algorithm are designed for use in a multiresolution transform domain. These are often presented in an implementation using one type of multiresolution transform. The basic principles however, are general enough to allow the techniques to be used with any multiresolution image decomposition. That is why it was chosen to explain the algorithms for a general image transformation and refer to the multiresolution image decomposition transform coefficients as 'transform coefficients' or just simply 'coefficients'. Examples of possible image transforms are the Discrete Wavelet Transform (DWT) (Donoho, 1995), the Curvelet Transform (Starck, 2002), the Shearlet transform (Easley, 2006), the dual-tree complex wavelet transform (Kingsbury, 2001; Selesnick, 2005), Steerable pyramids (Simoncelli, 1992), ... Since this chapter deals with the difficulties of handling correlated noise, an investigation into the differences between the mentioned image transforms is beyond the scope of this work.

5.1 Noise Whitening

A convenient way to deal with the problem of correlated noise is to start from existing denoising techniques. Most existing denoising techniques assume that noise is uncorrelated. One way to be able to reuse existing denoising techniques is to use a prewhitening transform that decorrelates the image noise. Arranging the noise measurements in a vector **f**, the decorrelating transform matrix **W** needs to satisfy the whiteness property:

$$\begin{aligned} &E\left[f.f^{*}\right]=C \\ &E[W.f.f^{*}.W^{*}]=I \\ &W.C.W^{*}=I \\ &C=(W^{*}.W)^{-1} \\ &W=C^{-1/2} \end{aligned} \tag{17}$$

The result from (17) shows that the decorrelating transform can be calculated as the matrix square root of the matrix inverse of the covariance matrix **C**. The covariance matrix **C** and the autocorrelation function r(n) are closely related through their respective definitions (again, a zero mean noise process f() is assumed):

$$\begin{aligned} &r(n)=E[f(m)\overline{f(n+m)}] \\ &C=E[f.f^{t}] \end{aligned} \tag{18}$$

The ability to whiten transform coefficients with respect to the noise indicates it is possible to use the techniques for covariance estimation from section 4 and the whitening transformation in combination white noise denoising algorithms.

In practice, for many types of correlated noise, it is not necessary to take correlation between samples across the entire image into account. In fact, this would make the whitening procedure (17) intractable. When considering a size N for the neighbourhood vector **f**, it means that the noise covariance matrix **C** is of size N^2. That is why in practice, the neighbourhood size N is kept low, often just 9=3x3 or 25=5x5. The approximation means that all correlation with pixels not included in the neighbourhood window is set to zero, but this is usually not problematic in practice.

The whitening operation in (17) can become quite computationally expensive, as a naïve implementation would mean vectorizing an image into a one neighbourhood vector per pixel, causing huge memory requirements and then one matrix multiplication per pixel. Since the interest is in decorrelating the center pixel in the neighbourhood vector from its neighbours, actually only the corresponding row **w** of the decorrelating transform matrix **W** is of interest. The image now just needs to be filtered by this linear filter **w**, which has an interpretation as a prewhitening linear filter. In fact, given the PSD $R(\omega)$, the prewhitening filter $w(\omega)$ can also be obtained as:

$$w(\omega) = \frac{1}{\max(\varepsilon, \sqrt{R(\omega)})} \tag{19}$$

Where ε is a small constant to enforce stability against small values of the PSD $R(\omega)$. Note that the prewhitening filter (19) can be different when compared to the one derived from the autocovariance matrix **C.** The definition of **C** does not assume stationarity of the noise, which means that e.g. the diagonal elements can differ from each other. In this case the PSD $R(\omega)$ does not exist and (19) and (17) are not equivalent. In this chapter, we avoided that situation by assuming noise stationarity.

5.2 Bayesian Least Squares – GSM

The BLS-GSM denoising algorithm, proposed in (Portilla, 2005), is a multiresolution minimum mean square error estimator under multivariate GSM prior, explained in section 4. This makes it very suitable for correlated noise. It starts from the very generic Bayesian Least Squares (or MMSE) estimate for a noise free vector of transform coefficients:

$$E[x \mid y] = \int_x x p_{x|y}(x \mid y) dx \tag{20}$$

The conditional probability can be quite hard to calculate, but since the algorithm works in the multiresolution transform domain, it is possible to use the GSM prior model to simplify the calculation:

$$E[x \mid y] = \int_x x \int_z p_{x|z,y}(x \mid z, y) p_{z|y}(z \mid y) dz dx$$
$$= \int_z E[x \mid z, y] p_{z|y}(z \mid y) dz \tag{21}$$

Under the assumption of (multivariate) Gaussian noise, the expectations in (21) reduce to the Wiener estimate:

$$E[x \mid z, y] = zC_u.\left(zC_u + C_w\right)^{-1} y \tag{22}$$

In order to evaluate (21), it is also necessary to estimate the conditional probability of the hidden GSM multiplier z on the observed coefficient vector **y** where C_w is estimated as explained in section 4 and $C_u = E[z]^{-1}\left[C_y - C_w\right]_+$. With knowledge of the hidden multiplier pdf p(z), as explained in section 4, this is possible:

$$p_{z|y}(z \mid y) = \frac{p(y \mid z) p(z)}{p(y)} = \frac{p(y \mid z) p(z)}{\int_z p(y \mid z) dz} \tag{23}$$

where $p(\mathbf{y} \mid z)$ is the multivariate Gaussian distribution, given by (14). Implementation of these equations for the subbands of a multiresolution decomposition constitutes a BLS-GSM denoising algorithm.

5.3 Probshrink

Probshrink was first developed in (Pizurica, 2006), for white noise. In a more recent work (Aelterman, 2008) it has been extended to 3D correlated noise for denoising noise volumes. In section 3, it was explained how natural images consist of edges and smooth areas, giving rise to the heavy-tailed transform coefficient distributions. This model led to the idea of classifying transform coefficients into two classes. Transform coefficients are either large, meaning they represent an edge and arose from the tail of the distribution, or they are small, meaning they represent a smooth area (or the noise floor) and arose from the main body of the distribution:

$$H_1 = \mid y \mid > T$$
$$H_0 = \mid y \mid \leq T \tag{24}$$

with T a threshold value. T is often chosen to comply with the noise floor, related to the noise variance. In fact, this noise variance is often simply equal to the noise variance in the image, because some wavelet transforms are unitary with respect to white Gaussian noise. Conditioned on these two classes, the MMSE estimator becomes:

$$E[x \mid y] = E[x \mid y, H_1] P_{H|y}(H_1 \mid y) + E[x \mid y, H_0] P_{H|y}(H_0 \mid y) \tag{25}$$

Probshrink handles the transform coefficients coefficient-wise. It does not handle neighbourhood vectors of transform coefficients, as BLS-GSM does. In fact, this is not a problem for the decimated wavelet transform, as one can prove that the noise in the wavelet domain is white when the input noise was white. For redundant image transforms, this is

not the case, and an approach as in section 4.3 would show that the noise is in fact correlated in the transform domain. A proper approach for correlated noise is then needed, using a whitening transform to whiten coefficients to obtain better results.

The calculation of (25) is facilitated through some simplifying assumptions: It is assumed that $E[x \mid y, H_0] = 0$ because the model suggests that small transform coefficients, corresponding to hypothesis H_0, are attributed to noise. Similarly, large wavelet coefficients are attributed to the signal with very high SNR, and since this corresponds to hypothesis H_1, the assumption $E[x \mid y, H_1] = y$ is made. This leads to the Probshrink denoising estimator:

$$E[x \mid y] = P_{H|y}(H_1 \mid y)y \tag{26}$$

The Probshrink estimator can be seen as a soft shrinkage wavelet denoising operator, where the shrinkage factor is determined as $P_{H|y}(H_1 \mid y)$ through the Bayesian interpretation (25). Evaluation is done by rewriting:

$$E[x \mid y] = \frac{p_{y|H}(y \mid H_1)P_H(H_1)}{p_{y|H}(y \mid H_1)P_H(H_1) + p_{y|H}(y \mid H_0)P_H(H_0)}y \tag{27}$$

Assuming additive noise (with distribution $p_w(w)$), this expression can be evaluated quite easily. With knowledge of the prior model $p_x(x)$, the conditional probabilities in (27) are written as in (28).

$$\begin{aligned} p_{y|H}(y \mid H_0) &= \int p_w(y-x)p_{x|H}(x \mid H_0)dx \\ p_{y|H}(y \mid H_1) &= \int p_w(y-x)p_{x|H}(x \mid H_1)dx \end{aligned} \tag{28}$$

Hence, the conditional probabilities of $p_{y|H}$ in (28) are the convolved prior conditional probabilities $p_{x|H}$ with the noise distribution p_w. This is illustrated in figure 13.

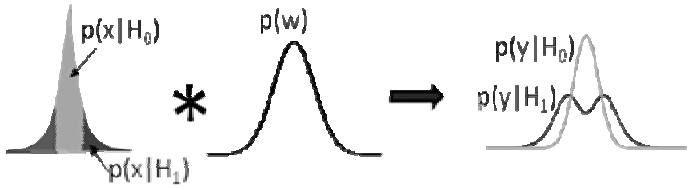


Fig. 13. Example of the estimation of the conditional probabilities $p_{y|H}$ from the prior probabilities $p_{x|H}$ and noise distribution p_w

Note how the conditional prior probability distributions $p_{x|H}$ can easily be calculated by imposing (24) and renormalizing the prior, also illustrated in figure 13. Again using (24), it is possible to estimate p_H through integrating this function over the threshold interval:

$$p_H(H_0) = \int_{-T}^{T} p_x(x)dx \qquad (29)$$
$$p_H(H_1) = 1 - p_H(H_0)$$

With knowledge of the exact prior model, it becomes possible to evaluate the Probshrink equations numerically (Pizurica, 2006), it is even possible to calculate (27) empirically, from the data (Pizurica, 2003) or analytically, for the Laplacian Prior (Pizurica, 2007). Together, these equations constitute the white noise Probshrink denoising method.

5.4 3D Probshrink for correlated noise

There exist several adaptations for the Probshrink estimator with the aim of handling correlated noise. In (Goossens, 2007; Goossens, 2009), such methods were proposed, which use a prewhitening filter before the classification step (24). In (Aelterman, 2008) the approach was extended to 3D, with the practical application of MRI denoising in mind. As both approaches are extensions of the Probshrink estimator, both approaches make use of multiresolution image decompositions (in this case, the dualtree complex wavelet transform (Kingsbury, 2001)). Instead of the classification (24), (30) is used:

$$H_1 = \left\| W.\overline{x_n} \right\| > T \qquad (30)$$
$$H_0 = \left\| W.\overline{x_n} \right\| \le T$$

With **W** the whitening transform matrix (8) and $\overline{y_n}$ the neighborhood coefficient vector of the coefficient at spatial location **n.** This will ensure that only coefficients are selected that are sufficiently distant from the noise in the succeeding Probshrink steps.

Multiresolution decompositions do not decorrelate the image features through the scales, a large transform coefficient in one scale increases the possibility of finding another in the subsequent scales. These interscale dependencies have been studied thoroughly for the wavelet transform. In (Mallat, 1998), it is shown that the wavelet coefficient magnitude increases with increasing scale (i.e. decreases with increasing spatial frequency) for natural, sufficiently regular signals, like natural images. Improvements to the Probshrink approach are made by exploiting this. In (Goossens, 2007), this is exploiting by fitting a Hidden Markov Tree model to the interscale dependencies of complex wavelet coefficients. In (Aelterman, 2008), this has been exploited by changing (30) by its interscale product:

$$H_1 = \left\| W_{j+1}.\overline{y_{j,n}} \right\| . \left\| W_j.\overline{y_{j,n}} \right\| > T \qquad (31)$$
$$H_0 = \left\| W_{j+1}.\overline{y_{j,n}} \right\| . \left\| W_j.\overline{y_{j,n}} \right\| \le T$$

With j the scale indicator. This way, a coefficient is only selected as significant coefficient if both it is large, as well as its whitened 'parent' coefficient, in compliance with the interscale dependency model.
These adaptations will make sure that only significant (with respect to the noise covariance) transform coefficients are selected, which will in the end lead to a better Probshrink denoising.

5.5 Non Local Means

The Non Local Means Algorithm (NLMS) is an algorithm that tries to exploit the self-similarity (section 3.2) in images. An overview is given in (Buades, 2008). A very intuitive way to look at this algorithm is through its relationship with the simpler Bilateral Filtering scheme (Tomasi, 1998). The simplest image denoising algorithms apply some kind of linear filter. The Bilateral Filtering modifies this principle by making the filter data-adaptive:

$$\widehat{x}(n) = \sum_m h(m,n)c\big(y(m), y(n)\big)y(n) \tag{32}$$

In short, noise-free pixel estimate $\widehat{x}$ is a linear combination of noisy pixel values y with the weights depending on both the spatial position of the pixels (most of the time just the relative distance), and the actual value of the pixel. This way, through a clever choice of the weighing function, it is possible to reinforce statistical averaging of only similar pixel values, avoiding blurring of image edges.
The NLMS algorithm improves on this concept, by performing block matching on blocks (neighbourhoods) of pixels. Block matching uses the Euclidian distance between two vectors $\overline{y_n}$ and $\overline{y_m}$, containing the neighbourhood pixel values of a region around a pixels $y(n)$ and $y(m)$. It essentially compares image patches instead of pixels. This allows to find similar pixel values more accurately:

$$\widehat{x}(n) = k_n^{-1} \sum_m \exp\left(- \frac{\| \overline{y_n} - \overline{y_m} \|^2}{2h^2} \right) y(n) \tag{33}$$

With k_n a normalization constant. An important difference with the bilateral filter is that the NLMS theory allows to look beyond the linear filter mask for similar image patches. The block matching concept will ensure that similar pixels are found from a similar textured patch or image structure, even when the blocks come from the other side of the image. In practice this search space size condition will often be relaxed for computational reasons. The NLMS algorithm can hence be seen as an algorithm that exploits the self-similarity in an image.
The block matching approach means that the weights are calculated on neighbourhood vectors. From this it is conceptually a small step to do it on whitened neighbourhood vectors, using the noise whitening procedures from section 5.1. This decreases the influence of spatial noise correlation on the distance/similarity measure. The Euclidian distance in (33) is transformed into what is known as the Mahalanobis distance. This was one improvement over the classical NLMS scheme proposed in (Goossens, 2008).

$$\hat{x}(n) = k_n^{-1} \sum_m \exp\left(- \frac{\left(\overline{y_n} - \overline{y_m}\right)^t C_w^{-1} \left(\overline{y_n} - \overline{y_m}\right)}{2h^2} \right) y(n) \tag{34}$$

Using the Mahalanobis distance, the measure will actually take the correlation structure of the noise into account when calculating the distance between two image patches.

6. Results

The next few figures show a comparison of the two classes of algorithms, with different types of noise. Peak signal to noise ratio (PSNR) comparisons are given in table 1.
Figure 14 shows the result of white noise denoising on the Einstein image. For white noise it can be seen that the denoising algorithms perform similar with only a relatively small difference in PSNR. A visible difference is that the NLMS algorithm suffers less from the typical wavelet artifacts (small very local oscillations in image intensity).

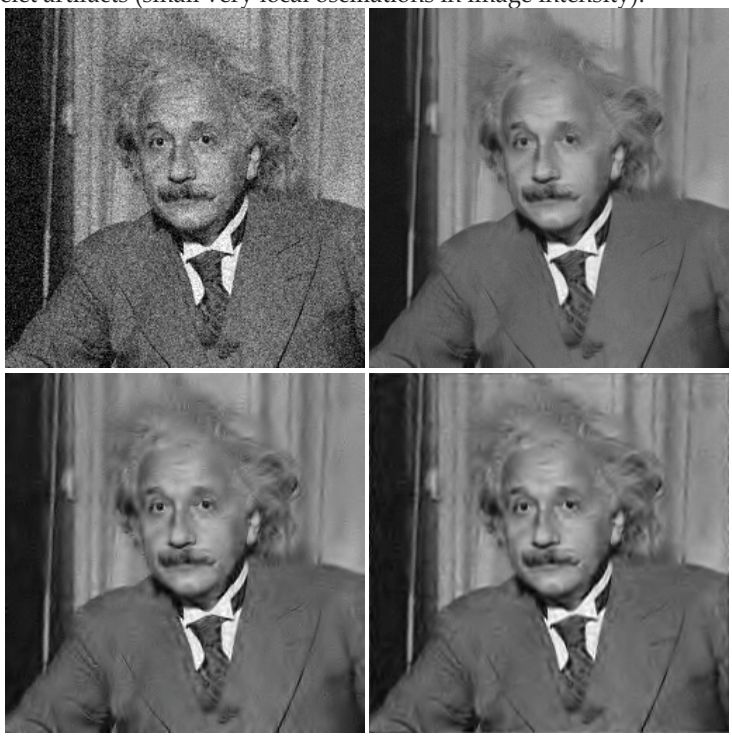


Fig. 14. Denoising results of Einstein image corrupted by simulated white noise (top left): by NLMS denoising (top right), by BLS-GSM denoising (bottom left), by Probshrink for white noise (bottom right)

In a second experiment, thermal noise was simulated on the House image. This is correlated noise, which behaves as stripes. It can be seen that an algorithm for correlated noise denoising has some advantage in PSNR. A qualitative comparison can be made in figure 15. Clearly, the Probshrink algorithm, which has very good results for white noise, has

problems with properly discerning image details from the noise structure. Even though the correlated noise algorithms succeed in preserving the horizontal brick structure of the house to some degree, the white noise Probshrink algorithm completely suppresses those details while preserving the vertical stripes noise pattern.

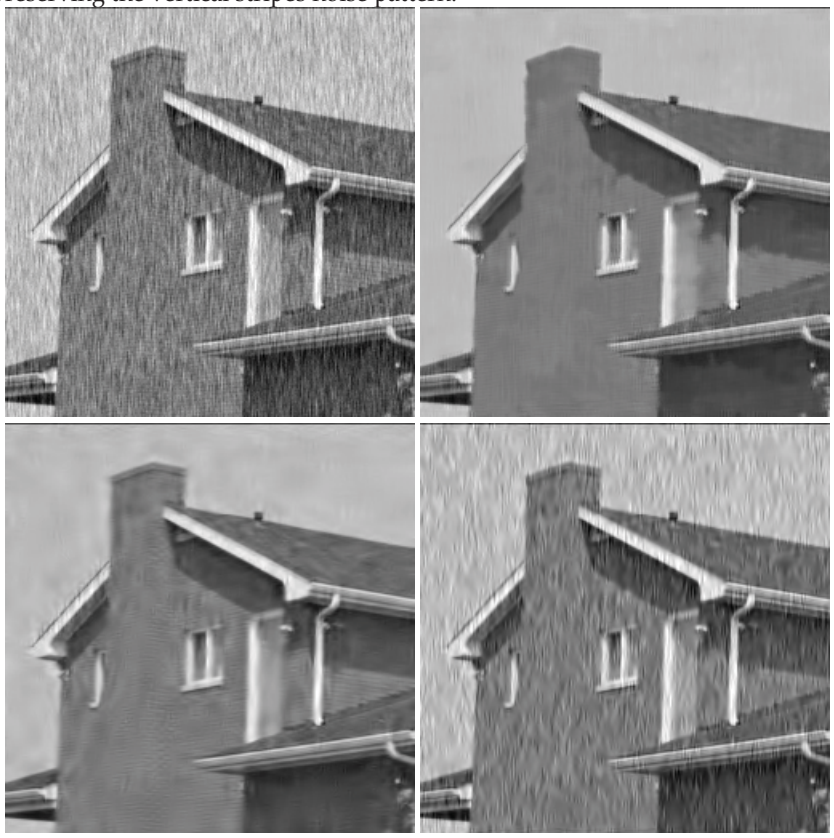


Fig. 15. Denoising results of House image corrupted by simulated thermal noise (top left): by NLMS denoising (top right), by BLS-GSM denoising (bottom left), by Probshrink for white noise (bottom right)

In a third denoising experiment, a fictional type of correlated noise was used, high frequency noise, combined with diagonal streaks. It is referred to as 'streak' noise in Table 1. The qualitative comparison can be seen in figure 16. As expected, the white noise denoising algorithm Probshrink does not succeed in suppressing the diagonal streak noise structures. It does succeed in suppressing the high frequency noise, because even for white noise the high frequency parts of the noisy image PSD typically have a very low SNR, because of the fact that natural images predominantly have energy in low frequency parts of the PSD, as explained in section 3.

Fig. 16. Denoising results of Goldhill image corrupted by heavily correlated streak noise (top left): by NLMS denoising (top right), by BLS-GSM denoising (bottom left), by Probshrink denoising for white noise (bottom right)

In a fourth denoising experiment, the Stonehenge image was used. It was treated as a color image, and used as input for a mosaicing/demosaicing experiment using the bilinear demosaicing algorithm. This results in low frequency noise structures. Then the red channel of the resulting color image was used as input for the denoising experiment. Again, it is visible that the white noise denoising algorithm Probshrink does not succeed in suppressing the noise artifacts, while the algorithms for correlated noise do. It is also visible that the BLS-GSM algorithm suffers from ringing near the top edge of the Stonehenge structure. This type of artifacts is common in wavelet-base denoising experiments and is a result from incorrectly suppressing the small coefficients that make up the edge in higher frequency scales, while keeping their respective counterparts in lower frequency scales.

Fig. 17. Denoising results of Stonehenge image corrupted by simulated red channel demosaicing noise (top left): by NLMS denoising (top right), by BLS-GSM denoising (bottom left), by Probshrink denoising for white noise (bottom right)

From the experiments, some conclusions can be made. White noise denoising algorithms, such as Probshrink, work well enough as long as the image is corrupted by white noise. It fails when presented with correlated noise. One reason is that the Donoho MAD estimator is often a very bad choice, leading to underestimated noise power (for low frequency noise) or severely overestimated noise power (for high frequency noise). Because of this failure of the MAD estimator, the choice was made to choose the noise variance parameter heuristically for the white noise Probshrink algorithm, in order to obtain the highest possible PSNR. It can be concluded from figures 14-17 and table 1, that for situations where image noise is correlated, a simple white noise denoising algorithm will not perform optimally and there is need for the techniques and ideas explained in this chapter.

	Noisy	ProbShrink	BLS-GSM	NLMS
White	22dB	29.22dB	29.76dB	29.88dB
Demosaicing	27.9dB	29.8dB	32.6dB	31.4dB
Thermal	24.5dB	26.0dB	31.6dB	31.5dB
Streaks	16.1dB	22.8dB	25.7dB	25.9dB

Table 1. PSNR table for the different denoising experiments

In a last experiment, we used the 3D dual tree complex wavelet denoising algorithm for MRI (Aelterman, 2008) to illustrate the denoising performance on practical MRI images. A qualitative comparison can be seen in figure 18.

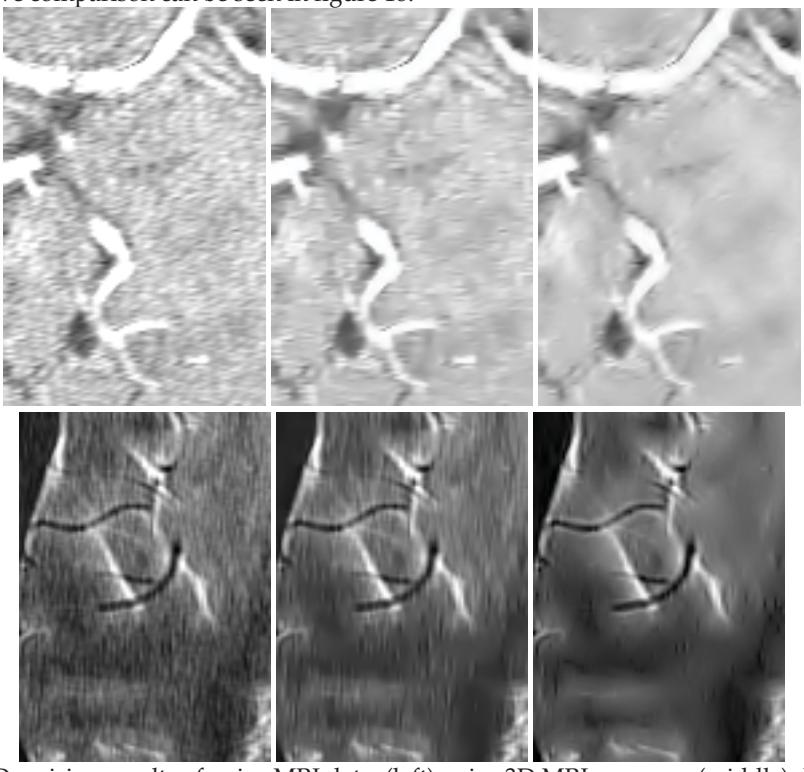


Fig. 18. Denoising results of noisy MRI data. (left) noisy 3D MRI sequence (middle) denoised by 2D per-slice Probshrink (right) denoised by 3D correlated noise Probshrink for MRI

7. Conclusion

From the results in the previous section, it is clear that one needs to make use of specialized denoising algorithms for situations in which one encounters correlated noise in images. The short overview in section 2 shows that there are many such situations in practice. Correlated noise manifests itself as stripes, blobs or other image structures that cannot be modelled as spatially independent. Several useful noise estimation techniques were presented that can be used when creating or adapting a white noise denoising algorithm for use with correlated noise. To illustrate this, some state-of-the-art techniques were explained and compared with techniques designed for white noise.

8. References

Aelterman, J.; Goossens, B.; Pizurica, A. & Philips, W. (2008) Removal of Correlated Rician Noise in Magnetic Resonance Imaging *Proceedings of European Signal Processing Conference (EUSIPCO, Lausanne, 2008*

Aelterman, J.; Goossens, B.; Pizurica, A. ; Philips, W. (2009) Locally Adaptive Complex Wavelet-Based Demosaicing for Color Filter Array Images *Proceedings of SPIE Electronic Imaging 2009, San Jose, CA*, Vol. 7248, no. 0J

Bayer, B. (1976) Color Imaging Array *US Patent 3,971,065*

Borel, C.; Cooke, B.; Laubscher, B. (1996) Partial Removal of Correlated noise in Thermal Imagery *Proceedings of SPIE*, Vol. 2759, 131

Buades, A., Coll B. & Morel J. M. (2005) Image Denoising by Non-Local Averaging, *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 25-28

Buades, A.; Coll, B & Morel, J.M (2008) Nonlocal Image and Movie Denoising *Int Journal on Computer vision* Vol. 76, pp. 123-139

Dabov, K.; Foi, A.; Katkovnik, V. & Egiazarian K. (2006) Image Denoising with Block-Matching and 3D Filtering, *Proc. SPIE Electronic Imaging: Algorithms and Systems V*, no. 6064A-30

Dabov, K.; Foi, A.; Katkovnik, V. & Egiazarian K. (2007) Image denoising by sparse 3D transform-domain collaborative filtering, *IEEE Trans. on Im. Processing*, vol. 16, no. 8

Donoho, D. & Johnstone, I. (1994) Adapting to Unknown Smoothness via Wavelet Shrinkage *Journal of the American Statistics Association*, Vol. 90

Donoho, D. L. (1995) De-Noising by Soft-Thresholding, *IEEE Transactions on Information Theory*, vol. 41, pp. 613-62.

Easley, G.; Labate, D.; Lim, Wang-Q, (2006) Sparse Directional Image Representation using the Discrete Shearlet Transform *Preprint submitted to Elsevier Preprint*

Elad, M.; Matalon, B.; & Zibulevsky, M. (2006) Image Denoising with Shrinkage and Redundant Representations *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* vol. 2, pp. 1924-1931

Field, D. (1987) Relations between the statistics of natural images and the response properties of cortical cells *J. Opt. Soc. Am.* A 4, p. 2379-2394

Goossens, B.; Pizurica, A. & Philips, W. (2007) Removal of Correlated Noise by Modeling Spatial Correlations and Interscale Dependencies in the Complex Wavelet Domain *Proceedings of International Conference on Image Processing (ICIP)* pp. 317-320

Goossens, B.; Luong, H., Pizurica, A. Pizurica & Philips, W. (2008) An Improved Non-Local Denoising Algorithm *Proceedings of international Workshop on Local and Non-Local Approximation in Image Processing*, Lausanne, 2008

Goossens, B.; Pizurica, A. & Philips W. (2009) Removal of correlated noise by modelling the signal of interest in the wavelet domain *IEEE Transactions on Image Processing* in press

Guerrero-Colon, J. ; Simoncelli, E. & Portilla, J. (2008) Image Denoising using Mixtures of Gaussian Scale Mixtures, *Proc. IEEE Int. Conf. on Image Processing (ICIP),* San Diego, 2008.

Hastie, Trevor; Tibshirani, Robert & Friedman, J. (2001) The Elements of Statistical Learning *New York: Springer* 8.5 The EM algorithm pp. 236–24

Kingsbury, N. G. (2001) Complex Wavelets for shift Invariant analysis and Filtering of Signals, *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234-253

Kwon, O.; Sohn, K. & Lee, C. (2003) Deinterlacing using Directional Interpolation and Motion Compensation *IEEE Transactions on Consumer Electronics,* vol. 49, no. 1

Malfait, M. & Roose, D. (1997) Wavelet-Based image denoising using a Markov random field a priori model. *IEEE Transactions on Image Processing*, vol. 6, no. 4, pp. 549-565

Mallat, S. (1989) A theory for multiresolution signal decomposition: the wavelet representation *IEEE Pat. Anal. Mach. Intell.,* Vol. 11, pp. 674-693

Mallat, S. (1998) A Wavelet Tour of Signal Processing, *Academic Press, 1998*, p. 174

Nowak, R. (1999) Wavelet-based Rician noise removal for Magnetic Resonance Imaging *Transactions on Image Processing,* vol. 10, no. 8, pp. 1408-1419

Pizurica, A.; Philips, W.; Lemahieu, I. & Acheroy, M. (2003) A Versatile Wavelet Domain Noise filtration Technique for Medical Imaging *IEEE Transactions on Medical Imaging,* vol. 22, no. 3, pp. 323-331

Pizurica, A. & Philips, W. (2006) Estimating the Probability of the Presence of Signal of Interest in Multiresolution Single- and Multiband Image Denoising *IEEE Transactions on Image Processing,* Vol. 15, No. 3, pp. 654-665

Pizurica, A. & Philips, W. (2007) Analysis of least squares estimators under Bernoulli-Laplacian priors *Twenty eighth Symposium on Information Theory in the Benelux* Enschede, The Netherlands, May 24-25 2007

Portilla, J.; Strela, V.; Wainwright, M.J. & Simoncelli, E.P. (2003) Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain *IEEE Transactions On Image Processing*, vol. 12, no. 11., pp. 1338-1351

Portilla, J. (2004) Full Blind Denoising through Noise Covariance Estimation using Gaussian Scale Mixtures in the Wavelet Domain, *Proc. IEEE Int. Conf. on Image Processing (ICIP),* pp. 1217-1220

Portilla, J. (2005) Image Restoration using Gaussian Scale Mixtures in Overcomplete Oriented Pyramids *SPIE's 50th Annual Meeting, Proc. of the SPIE,* vol. 5914, pp. 468-82

Romberg, J; Choi, H. & Baraniuk R. (2000) Bayesian Tree-Structured Image Modeling using Wavelet-domain Hidden Markov Models *IEEE Transactions on Image Processing*, vol. 10, no. 7

Ruderman, D (1994) The statistics of natural images *Network: Computation in Neural Systems*, Vol. 5, pp. 517-548

Selesnick, I.W.; Baraniuk, R.G. & Kingsbury, N.G. (2005) The Dual-Tree Complex Wavelet Transform, *IEEE Signal Processing Magazine*, pp 123-151

Simoncelli, E.; Freeman, W.; Adelson, E. & Heeger D. (1992) Shiftable Multi-Scale Transforms or, "What's Wrong with Orthonormal Wavelets" *IEEE Trans. Information Theory, Special Issue on Wavelets Vol. 38, No. 2, pp. 587-607*

Starck, J.-L; Candès, E. J. & Donoho, D. L. (2002) The Curvelet Transform for Image Denoising, *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670-684

Tomasi, C. & Manduchi, R. (1998) Bilateral Filtering for Gray and Color Images *Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay,India, 1998*

Wainwright, J. & Simoncelli, E. (2000) Scale Mixtures of Gaussians and the Statistics of Natural Images *Advances in Neural Information Processing Systems*, Vol. 12, pp. 855-861