CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

` Digital images play an important role both in daily life applications such as satellite television, magnetic resonance imaging, computer tomography as well as in areas of research and technology such as geographical information systems and remote sensing. Data sets collected by image sensors are generally contaminated by noise. Imperfect instruments, problems with the data acquisition process, and interfering natural phenomena can all degrade the data of interest. Furthermore, noise can be introduced by transmission errors and compression. Thus, denoising is often a necessary and the first step to be taken before the images data is analyzed. It is necessary to apply an efficient denoising technique to compensate for such data corruption. Removing noise from the noisy signal is a challenge for researchers because noise removal introduces artifacts and causes blurring of the images. There are many different approaches for noise reduction. Each approach has its assumptions, advantages, and limitations. Most of the natural images are assumed to have additive random noise which is modeled as a Gaussian distribution **[Moulin 1999]**. Speckle noise **[Guo 1994]** is observed in SAR images whereas Rician noise **[Nowak 1999]** affects Magnetic resonance images (MRI). This chapter reviews some of the significant work in this area of image denoising.

2.2 CLASSIFICATION OF DENOISING ALGORITHMS

In general, the denoising methods can be grouped as spatial domain and transform domain. The spatial domain methods can again be sub grouped as linear and nonlinear denoising methods. In transform domain methods, wavelet transform gives a performance superior in image denoising due to its four basic properties:

- **Energy Conservation-** The wavelet transform is an orthogonal transform.
- **Energy Compaction-** There are large numbers of small magnitude wavelet coefficients. Most of the energy is concentrated in the trend.
- **Two Populations-** The larger wavelet coefficients are clustered around sharp transition regions (edges in images). Smaller wavelet coefficients reside in smoother regions.
- **Clustering-** Large magnitude wavelet coefficients tend to have some large magnitude coefficients located near them.

With WT gaining popularity in the last two decades, various algorithms for denoising in wavelet domain were introduced **[Balster 2003]**, **[Baraniuk 1999]**, **[Biao 2001]**, **[Cai 2001]**, **[Chang 2000]**. The focus was shifted from the Spatial and Fourier domain to the WT domain. Ever since Donoho's wavelet based thresholding approach **[Dohono 1995]** was published, there was a surge in the denoising papers being published. Although Donoho's concept **[Dohono 1994]** was not revolutionary, his methods did not require tracking or correlation of the wavelet maxima and minima across the different scales as proposed by Mallat **[Mallat 1992]**. Thus, there was a renewed interest in wavelet based denoising techniques since Donoho demonstrated a simple approach to a difficult problem. Researchers published different ways to compute the parameters for the thresholding of wavelet coefficients **[Antonini 1992]** and **[Azalini 2005]**. Data adaptive thresholds **[Fodor 2001]** were introduced to achieve optimum value of threshold. Later efforts found that substantial improvements in perceptual quality could be obtained by translation invariant methods based on thresholding of an undecimated wavelet transform **[Donoho 1995]**. These thresholding techniques were applied to the non-orthogonal wavelet coefficients to reduce artifacts. Multiwavelets were also used to achieve similar results. Probabilistic models using the statistical properties of the wavelet coefficient seemed to outperform the thresholding techniques and gained ground. Recently, much effort has been devoted to Bayesian denoising in wavelet domain **[Ray 2003]**. Hidden Markov Models and Gaussian Scale Mixtures have also become popular and more research in this area is going on. Tree Structures ordering the wavelet coefficients based on their magnitude, scale and spatial location have been proposed by Scharcanski **[Scharcanski 2002]**. Data adaptive transforms such as Independent Component Analysis (ICA) have been explored for sparse shrinkage **[Moresan 2003]**. The trend continues to focus on using different statistical models to model the statistical properties of the wavelet coefficients and its neighbors **[Portilla 2001]** and **[Portilla 2003]**. Future trend will be towards finding more neighboring pixel adaptive shrinkage functions to get better results. There are two basic approaches to image denoising, spatial filtering methods and transform domain filtering methods.

2.2.1 Spatial Filtering

A traditional way to remove noise from image data is to employ spatial filters. The pixel intensity information is manipulated in same resolution and same domain. Spatial filters can be further classified into non-linear and linear filters.

2.2.1.1 Non-Linear Filtering

With non-linear filters, the noise is removed without any attempts to explicitly identify it. Spatial filters employ a low pass filtering on groups of pixels with the assumption that the noise occupies the higher region of frequency spectrum. Generally spatial filters remove noise to a reasonable extent but at the cost of blurring images which in turn makes the edges in pictures invisible. In recent years, a variety of nonlinear median type filters such as rank conditioned rank selection **[Hardie 1994]**, weighted median **[Yang 1995]**, mode filtering **[Evans 1995]** and relaxed median **[Hamza 1999]** have been developed to overcome this drawback. The recursive algorithm based on Least mean square (LMS) has been proposed by Chen **[Chen 2001]** which give better results than non-recursive methods.

2.2.1.2 Linear Filtering

A mean filter is the optimal linear filter for Gaussian noise in the sense of mean square error. Linear filters too tend to blur sharp edges, destroy lines and other fine image details, and perform poorly in the presence of signal-dependent noise. The best linear filter is Weiner filter. The Wiener filtering method **[Jain1989]** requires the information about the spectra of the noise and the original signal and it works well only if the underlying signal is smooth. Wiener method implements spatial smoothing and its model complexity control correspond to choosing the window size. To overcome the weakness of the Wiener filtering, Donoho and Johnstone proposed the wavelet based denoising scheme in **[Donoho 1995]** and **[Donoho 2001]**.

2.2.2 Transform Domain Filtering

The transform domain filtering methods can be subdivided according to the choice of the basis functions. The global basis functions have been used in Fourier domain which would not give good results because image is stochastic process. The wavelet transform used local basis function which would give better results than frequency based methods. The spatial frequency and wavelet based methods are explained below.

2.2.2.1 Spatial-Frequency Filtering

The frequency domain methods have been assumed that information of signal has been laid in low frequency and noise in high frequency. Noise has been reduced by low pass filtering using Fast Fourier Transform (FFT). In frequency smoothing methods **[Jain 1989]**, the removal of the noise is achieved by designing a frequency domain filter and adapting a cut-off frequency when the noise components are decorrelated from the useful signal in the frequency domain. These methods are time consuming and depend on the cut-off frequency and the filter function behavior. Furthermore, they may produce artificial frequencies in the processed image. To overcome these drawbacks, the wavelet transform is used.

2.2.2.2 Wavelet transform domain

Filtering operations in the wavelet domain can be subdivided into linear and nonlinear methods. Linear filters such as Wiener filter in the wavelet domain yield optimal results when the signal corruption can be modeled as a Gaussian process and the accuracy criterion is the mean square error (MSE) **[Strela 2000]** and **[Choi 1998]**. However, designing a filter based on this assumption frequently results in a filtered image that is more visually displeasing than the original noisy signal, even though the filtering operation successfully reduces the MSE. In **[Zhang 2000]** a wavelet-domain spatially adaptive FIR Wiener filtering for image denoising is proposed where wiener filtering is performed only within each scale and intrascale filtering is not allowed. The most investigated domain in denoising using Wavelet Transform is the non-linear coefficient thresholding based methods. The procedure exploits sparsity property of the wavelet transform and the fact that the WT maps white noise in the signal domain to white noise in the transform domain. Thus, while signal energy becomes more concentrated into fewer coefficients in the transform domain, noise energy does not. It is this important principle that enables the separation of signal from noise. The procedure in which small coefficients are removed while others are left untouched is called hard thresholding **[Donoho 1995]**. But the method generates spurious blips, better known as artifacts, in the images as a result of unsuccessful attempts of removing moderately large noise coefficients. To overcome the demerits of hard thresholding, wavelet transform using soft thresholding was also introduced in **[Donoho 1995]**. In this scheme, coefficients above the threshold are shrunk by the absolute value of the threshold itself. Similar to soft thresholding, other techniques of applying thresholds are semi-soft thresholding and Garrote thresholding **[Fodor 2001]**. Most of the wavelet shrinkage literature is based on methods for choosing the optimal threshold which can be adaptive or non-adaptive to the image. The wavelet based methods can be grouped in three types- Universal threshold methods, Bayesian methods, translation invariant methods.

2.2.2.2.1 Universal Threshold Methods

Donoho and Johnstone **[Donoho 1994]** and Donoho **[Donoho 1995]** proposed a universal threshold which is also called VISU Shrink. VISU Shrink is non-adaptive threshold, which depends only on number of data points. It has asymptotic equivalence suggesting best performance in terms of MSE when the number of pixels reaches infinity. VISU Shrink is known to yield overly smoothed images because its threshold choice can be unwarrantedly large due to its dependence on the number of pixels in the image.

Donoho and Johnstone **[Donoho 1995]** extended the ideas behind the VISU Shrink to a level-dependent shrink and SURE Shrink. In level dependent shrink, the threshold value depends on number of data in each level. Similarly in SURE Shrink, a threshold that minimizes Stein's unbiased estimate of risk is applied to each decomposition scale unless it is determined that the coefficients at the level in question are negligible, in which case a VISU Shrink is applied.

2.2.2.2.2 Bayesian Methods

In Bayesian shrinkage methods, a prior is placed on all of the wavelet coefficients of the target signal. Chipman et al. **[Chipman 1997]** assumed that this prior is a mixture of two Gaussian distributions. They had minimized the Bayes Risk Estimator function assuming Generalized Gaussian prior and thus yielding data adaptive threshold. Most of the times, Bayes Shrink outperforms VISU Shrink and SURE Shrink. A problem with these methods and many other Bayesian thresholding approaches is the difficulty in estimating the hyperparameters. Abramovich et al. **[Abramovich 1998]** provide a theorem for choosing the hyperparameters based on prior knowledge of the regularity properties of the target function. However, they indicate that this may be a "daunting prospect" and provide standard choices for the hyperparameters. Figueiredo and Nowak **[Figueiredo 2001]** had solved the problem of estimating hyperparameters by introducing a simpler Bayesian model, the amplitude-scale-invariant Bayes estimator (ABE), which has no hyper parameters.

Another difficulty of some of these Bayesian models is the assumption of a single Gaussian component for the entire image. To assume that one particular density will be sufficient to model the density of image coefficients in all types of images may be too restrictive a model. Number of researchers has developed homogeneous local probability models for images in the wavelet domain. Specifically, the marginal distributions of wavelet coefficients are highly kurtotic, and usually have a marked peak at zero and heavy tails. Crouse et al. **[Crouse 1998]** used Gaussian and non-Gaussian mixture densities to model the priors on the wavelet coefficients. The Gaussian mixture model (GMM) **[Chipman 1997]** and the generalized Gaussian distribution (GGD) **[Moulin 1999]** are commonly used to model the wavelet coefficients distribution. Although GGD is more accurate, GMM is simpler to use. Mihcak proposed a methodology in which the wavelet coefficients are assumed to be conditionally independent zero-mean Gaussian random variables, with variances modeled as identically distributed, highly correlated random variables **[Mihcak 1999]**. An approximate Maximum A Posteriori (MAP) estimator is used to estimate marginal prior distribution of wavelet coefficient variances. Simoncelli and Adelson **[Simoncelli 1996]** used a two parameter generalized Laplacian distribution for the wavelet coefficients of the image, which is estimated from the noisy observations. Chang et al. **[Chang 2000b]** proposed the use of adaptive wavelet thresholding for image denoising by modeling the wavelet coefficients as a generalized Gaussian random variable, whose parameters are estimated locally (i.e., within a given neighborhood). Similarly, Cai et al. **[Cai 2001a, 2001b]** proposed threshold function which had incorporated neighboring information for one-dimensional signal.

In addition to restricting the shape of distribution, most of these Bayesian approaches assume independence of the wavelet coefficients. Other Bayesian thresholding methods have been developed that attempt to correct this deficiency. These approaches focus on exploiting the multiresolution properties of Wavelet Transform. These techniques identify close correlation of signal at different resolutions by observing the signal across multiple resolutions. They produce excellent output but is computationally much more complex and expensive. The modeling involves creating tree structure of wavelet coefficients with every level in the tree representing each scale of transformation and nodes representing the wavelet coefficients. This approach is adopted by Baraniuk **[Baraniuk 1999]**. The optimal tree approximation displays a hierarchical interpretation of wavelet decomposition. WT of singularities have large wavelet coefficients that persist along the branches of tree. Thus if a wavelet coefficient has strong presence at particular node then in case of it being signal, its presence should be more pronounced at its parent nodes. If it is noisy coefficient, for instance spurious blip, then such consistent presence will be missing. Lu et al. tracked wavelet local maxima in scale space by using a tree structure [**Lu 1992]**. Other denoising method based on wavelet coefficient trees is proposed by Mallat **[Mallat 1992a]**. Hidden Markov Models (HMMs) are used to nodes (significant or not) of the wavelet coefficients. The correlation between coefficients at same scale but residing in a close neighborhood are modeled by Hidden Markov Chain (HMC) model where as the correlation between coefficients across the chain is modeled by Hidden Markov Trees (HMT). Once the correlation is captured by HMM, Expectation Maximization (EM) is used to estimate the required parameters and then denoised signal is estimated using well known MAP estimator. Portilla described a model in which each neighborhood of wavelet coefficients is described as a Gaussian scale mixture (GSM), which is a product of a Gaussian random vector, and an independent hidden random scalar multiplier **[Portilla 2003]**. Strela et al. **[Strela 2000b]** described the joint densities of clusters of wavelet coefficients as a Gaussian scale mixture, and developed a maximum likelihood solution for estimating relevant wavelet coefficients from the noisy observations. The disadvantage of HMT is the computational burden of the training stage. In order to overcome this computational problem, a simplified HMT (uHMT) was proposed **[Romberg 2001]**. Another approach that uses a Markov random field [MRF] model for wavelet coefficients was proposed by Jansen and Bulthel **[Jansen 2001]**. Markov Random Field (MRF) **[Malfait 1997]** models are more efficient to capture intrascale correlations, whereas Hidden Markov Model (HMM) **[Romberg 2001]** is efficient in capturing inter-scale dependencies.

Another Bayesian approach that tries to capture the inter-scale dependency of wavelet coefficients is given by Sendur and Selesnick **[Sendur 2002a, 2002b]**. They proposed a bivariate probability density function (pdf) for parent/child pairs to model their dependence. While these models give improved estimates by modeling the dependence structure of wavelet coefficients, they still require the signal component to be specified. The methods explained above used DWT which have orthogonal basis functions.

2.2.2.2.3 Non-orthogonal Wavelet Transforms

Undecimated Wavelet Transform (UDWT) has also been used for decomposing the signal to provide visually better solution. Since UDWT is shift invariant it avoids visual artifacts such as pseudo-Gibbs phenomenon. Though the improvement in results is much higher, use of UDWT adds a large overhead of computations thus making it less feasible. Lang extended the concept of normal hard/soft thresholding to Shift Invariant Discrete Wavelet Transform **[Lang 1995]**. The Shift Invariant Wavelet Packet Decomposition (SIWPD) is exploited to obtain number of basis functions **[Cohen 1999]**. The Best Basis Function was found by Minimum Description Length principle which yielded smallest code length required for description of the given data. Then, thresholding was applied to denoise the data.

In addition to UDWT, use of Multiwavelets is explored which further enhances the performance but further increases the computation complexity. The Multiwavelets are obtained by applying more than one mother function (scaling function) to available dataset. Multiwavelets possess properties such as short support, symmetry, and the most importantly higher order of vanishing moments. This combination of shift invariance and Multiwavelets is implemented by Bui and Strela which give superior results for the Lena image in context of MSE **[Bui 1998]** and **[Strela 1999]**.

The Independent Component Analysis (ICA) has gained wide spread attention. The ICA method was successfully implemented for denoising Non-Gaussian data **[Jung 2001]** and **[Hyvarinen 1998]**. One exceptional merit of using ICA is its assumption of signal to be Non-Gaussian, which helps to denoise images with Non-Gaussian as well as Gaussian distribution. The drawbacks of ICA based methods are the computational cost because it uses a sliding window and it requires sample of noise free data or at least two image frames of the same scene. In some applications, it might be difficult to obtain the noise free training data.

2.3 SUMMARY

This chapter discusses the literature available in this area. Many of the denoising methods assume the noise model to be Gaussian. In reality, this assumption may not always hold true due to the varied nature and sources of noise. An ideal denoising procedure requires a priori knowledge of the noise, whereas a practical procedure may not have the required information about the variance of the noise or the noise model. Thus, most of the algorithms assume known variance of the noise and the noise model to compare the performance with different algorithms. Gaussian Noise with different variance values is added in the natural images to test the performance of the algorithms. Not all researchers use high value of variance to test the performance of the algorithm when the noise is comparable to the signal strength. Wavelet Transform is the best suited for performance because of its properties like sparsity, multiresolution and multiscale nature. In addition to performance, issues of computational complexity must also be considered. Thresholding techniques used with the Discrete Wavelet Transform are the simplest to implement. Non-orthogonal wavelets such as UDWT and Multiwavelets improve the performance at the expense of a large overhead in their computation. HMM based methods seem to be promising but are complex.

Several papers did not specify the wavelet used neither the level of decomposition of the wavelet transform was mentioned **[Nason 2002]**. It is expected that the future research will focus on building robust statistical models of wavelet coefficients based on their intra scale and inter scale correlations. Such models can be effectively used for image denoising and compression. All the methods mentioned above require a noise estimate, which may be difficult to obtain in practical applications. There are various methods to estimate noise variance. The methods are explained in next chapter.