CHAPTER 6
PROPOSED METHODOLOGY

This chapter presents algorithms and working of the proposed algorithm.

6.1 Proposed Methodology

The proposed algorithm presents the design and implementation of a PCA and Wavelet based image denoising technique. In this method, the input test image is subjected to additive white Gaussian noise having different noise variance levels. The noisy image is decomposed into different blocks. Then the calculation of principal components and their vectors for all the components is being done. As noise is having some common characteristics, the main task is search for the common and similar vectors in all the segments. Then the next job is to design and develop average vector from the entire common vector and estimate the components in the direction of average common vector for all the segments. After getting the common vector, covariance matrix is designed and the calculation of Eigen vectors and Eigen values is done. Apply the PCA transform and remove the above calculated components from the respective segments and deriving the new dimensional data set [6].

The key idea of Principal Component Analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the Principal Components (PCs), which are uncorrelated, and which are ordered so that the first few components retain most of the variations present in all of the original variables. Then calculate the variance of the above calculated components and finding the optimum threshold by using different wavelet shrinkage rule of threshold estimation using wavelet domain filtration with variance estimated. Here in this method ProbShrink rule is used for threshold estimation. The basic idea of this method is estimating the probability that a given coefficient contains a significant noise-free component, which is known as “signal of interest.”
For actual denoising, the simple shrinkage rule is that where empirical wavelet coefficients are multiplied with the probability of containing a significant noise-free component.

The heuristics behind this rule is to define a “softer” and yet simple alternative to the classical thresholding functions while satisfying the two main requirements: The shrinkage factor is always between zero and one and the coefficients that are more likely to represent the noise are heavily shrunk.

6.2 Proposed Algorithm

The proposed Algorithm can be described in following steps:

Step 1: Partition the noisy image into N X N equal blocks.

Step 2: Apply Algorithm and calculate the Principal Components & their Vectors for all the segments.

Step 3: Search for the common and similar vectors in all segments.

Step 4: Create the average vector from all common vectors.

Step 5: Estimate the components in the direction of average common vector for all segments.

Step 6: Calculate the Eigen vectors and Eigen values of the covariance matrix.

Step 7: Remove the above calculated components from the respective segments and deriving the new dimensional data set.

Step 8: Calculate the variance of the calculated components.

Step 9: Calculate the optimum threshold by using ProbShrink rule of threshold estimation using wavelet domain filtration with variance estimated in step 7.

Step 10: Calculate PSNR, MSE and Execution Time for evaluation of the algorithm.
6.3 Block Diagram of the Proposed Image Denoising Technique

Addition of Noise → PCA Transform → Thresholding → Inverse PCA Transform → Variance Calculation

Denoised Image → Optimum Thresholding → Wavelet Transform

Figure 6.1 Block Diagram of the Proposed Method
In this proposed method PCA as well as Wavelet Transform is used for denoising. With the rapid development of modern digital imaging devices and their increasingly wide applications in our daily life, to overcome the problem with conventional wavelet thresholding technique here PCA incorporated with wavelet transform is used. Wavelets are an efficient and practical way to retain edges and image information at multiple spatial scales. Image features at a given scale, such as houses or roads, can be directly enhanced by filtering the wavelet coefficients. Wavelets may be a more useful image representation than pixels.

Hence, in this algorithm PCA dimensionality reduction of wavelet coefficients is done in order to maximize edge information in the reduced dimensionality set of images. The wavelet transform will take place spatially over each image band, while the PCA transform will take place spectrally over the set of images. Thus, the two transforms operate over different domains. Still, PCA over a complete set of wavelet and approximation coefficients will result in exactly the same Eigen spectra as PCA over the pixels [81].

In this method, the noisy image is applied for principal component analysis (PCA). The output from PCA is the Eigen image and the eigenvectors. Apply soft thresholding on a PCA component, the number of them can be selected by the user, and the reconstruction quality in the inverse PCA (IPCA) depends on that number. Estimation of components in the direction of average common vector for all segments. The Eigen images and the eigenvectors are entropy coded. Then the wavelet transform (WT) is applied to that residual. Then apply thresholding on wavelet coefficient and the reconstruction quality in the Inverse Wavelet Transform (IWT) and finally get the output denoised image[83].

6.4. Mathematical Modeling of the Proposed Methodology

Principal component analysis (PCA) is a technique used for compression and classification of data. The purpose is to reduce the dimensionality for a data set by finding a new set of variables, smaller than the original set of variables and retains most of the sample’s information. By information it is meant that variation which is present in the sample, given by the correlations between the original variables. The new set of variables is known as principal
components (PCA), is uncorrelated, and is ordered by the fraction of the total information each retains.

### 6.4.1 Estimation of Principal Components

Let a sample of $n$ observations on vector of $p$ variables

$$X = (x_1, x_2, ..., x_p)$$

The first principal component of the sample by the linear transformation

$$z_1 \equiv a_1^T x = \sum_{i=1}^{p} a_{1i} x_i$$

Eq. 6.1

Where the vector $a_1 = (a_{11}, a_{21}, ..., a_{p1})$ is chosen such that $\text{VAR}[z_1]$ is maximum [38]

The $k^{th}$ Principal Component of the sample by the linear transformation is

$$z_k \equiv a_k^T x \quad k = 1, ..., p$$

Eq. 6.2

Where the vector $a_k = (a_{1k}, a_{2k}, ..., a_{pk})$

Is chosen such that $\text{var}[z_k]$ is maximum

Subject to $\text{COV}[z_k, z_l] = 0$ for $k > l \geq 1$

Eq. 6.3

and to $a_k^T a_k = 1$

### 6.4.2 Estimation of coefficient vectors $a_k$

For finding coefficient of $a_1$

$$\text{var}[z_1] = \langle z_1^2 \rangle - \langle z_1 \rangle^2$$

Eq. 6.4

$$= \sum_{i,j=1}^{p} a_{1i} a_{1j} \langle x_i x_j \rangle - \sum_{i,j=1}^{p} a_{1i} a_{1j} \langle x_i \rangle \langle x_j \rangle$$

Eq. 6.5

$$= \sum_{i,j=1}^{p} a_{1i} a_{1j} S_{ij}$$

Eq. 6.6

Where $S_{ij} \equiv \sigma_{x_i x_j} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$

$$= a_{1i}^2 S_{x_i}$$

Eq. 6.7
6.4.3 Estimation of Eigen Vector

$S$ is the covariance matrix for the variables $X=(x_1, x_2, ..., x_p)$

To find $a_1$ maximize $\text{var}[z_1]$ subject to $a_1^T a_1 = 1$

Let $\lambda$ be a Lagrange multiplier,

Then maximize $a_1^T S a_1 - \lambda (a_1^T a_1 - 1)$ \hspace{1cm} Eq. 6.8

By differentiating... $S a_1 - \lambda a_1 = 0$ \hspace{1cm} Eq. 6.9

$\Rightarrow (S - \lambda I_p) a_1 = 0$ \hspace{1cm} Eq. 6.10

Therefore $a_1$ is an eigenvector of $S$ corresponding to eigen value $\lambda \equiv \lambda_1$

Again maximize $\text{var}[z_1]$

$\text{Var}[z_1] = a_1^T S a_1 = a_1^T \lambda_1 a_1 = \lambda_1$ Eq. 6.11

So $\lambda_1$ is the largest eigen value of $S$

The first PC $z_1$ retains the greatest amount of variation in the sample.

Calculation of next coefficient vector $a_2$ maximize $\text{var}[z_2]$

Subject to $\text{COV}[z_2, z_1] = 0$

And to $a_2^T a_2 = 1$

First note that

$\text{COV}[z_2, z_1] = a_1^T S a_2 = \lambda_1 a_1^T a_2$ \hspace{1cm} Eq. 6.12

Then let $\lambda$ and $\phi$ are Lagrange’s multipliers and then maximize

$a_2^T S a_2 - \lambda (a_2^T a_2 - 1) - \phi a_2^T a_1$ \hspace{1cm} Eq. 6.13
It is observed that $a_2$ is also an eigen vector of $S$ where eigen value $\lambda \equiv \lambda_2$ is the second largest in the subset.

Therefore, $\text{Var}[z_k] = a_k^T S a_k = \lambda_k$ \hspace{1cm} Eq. 6.14

Thus $K_{th}$ largest eigen value of $S$ is actually the variance of $K_{th}$ principal component.

The $K_{th}$ PC $z_k$ retains the $K_{th}$ greatest fraction of the variation in the sample.

**6.4.4 Algebraic formulation of PCA**

Given a sample of $n$ observations on a vector of $p$ variables

$X = (x_1, x_2, \ldots, x_p)$

Now the vector of $p$ PC are

$z = (z_1, z_2, \ldots, z_p)$

Therefore $z = A^T x$ \hspace{1cm} Eq. 6.15

Where $A$ is orthogonal matrix of $p \times p$ whose $K_{th}$ column is the $K_{th}$ eigen vector $a_k$ of $S$.

Then $A = A^T S A$ is the covariant matrix of the PCs.

Being diagonal with elements $\Lambda_{ij} = \lambda_i \delta_{ij}$ \hspace{1cm} Eq. 6.16

Usage of PCA: Probability distribution

Probability distribution for sample PCs

If (i) the $n$ observations of $X$ in the sample are independent &

(ii) $X$ is drawn from an underlying population that

Follows a $p$-variate normal (Gaussian) distribution with known covariance matrix[38]

Then

$(n - 1)S \sim W_p(S, n - 1)$ \hspace{1cm} Eq. 6.17

Where $W_p$ is the Wishart distribution.
else utilize a bootstrap approximation

If
(1) \((n-1)S\) follows a Wishart distribution &
(2) the population eigen values \(\lambda'_k\) are all distinct

Then the following results hold as \(n\) approaches infinity

All the \(\lambda_k\) are independent of all \(a_k\)

\[
\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)
\]

& \(a_k\)

are jointly normally distributed

\[
< \lambda > = \lambda' \\
< a_k > = a'_k
\]

And

\[
\text{COV}[\lambda_k, \lambda_{k'}] = \begin{cases} \frac{2\lambda'^2_k}{n-1} k = k' \\ 0, \ k \neq k' \end{cases}
\]  

Eq. 6.18

\[
\text{COV}[a_{jk}, a_{jik'}] = \begin{cases} \frac{\lambda'_k \sum_{i=1}^{p} a_{ji}a'_{ji}}{(\lambda'_k - \lambda'_i)^2}, & k = k' \\ \frac{\lambda'_k \lambda'_i a_{ji}a'_{ji}}{(n-1)(\lambda'_k - \lambda'_i)^2}, & k \neq k' \end{cases}
\]  

Eq. 6.19

Thus knowing the \(k^{th}\) term retains the \(k^{th}\) greatest fraction of the variation so that one can approximate each observation by truncating the sum at the first where \(m\) is less than \(p\) PCs
Figure 6.2. Geometric picture of PC

\[ x_i \cong x_i^m = \sum_{k=1}^{m} z_{ik} a_k \]  
\hspace{1.5cm} \text{Eq. 6.20}

This reduces the dimensionality of the data

From \( p \) to \( m < p \) by approximating \( X \cong X^m = Z^mA^{mT} \)  
\hspace{1.5cm} \text{Eq. 6.21}

Where \( Z^m \) is the \( n \times m \) portion of \( Z \) and \( A^m \) is the \( p \times m \) potion of \( A \)

6.4.5 Estimation of Threshold

Here this method has used ProbShrink rule for threshold estimation. The basic idea of this method is estimating the probability that a given coefficient contains a significant noise-free component, which is known as “signal of interest.”

For actual denoising, the simple shrinkage rule is that where empirical wavelet coefficients are multiplied with the probability of containing a significant noise-free components [53].

The heuristics behind this rule is to define a “softer” and yet simple alternative to the classical thresholding functions while satisfying the two main requirements: The shrinkage factor is always between zero and one and the coefficients that are more likely to represent the noise are heavily shrunk.

The proposed approach removes the need for preliminary coefficient classifications and derives all the required probabilities analytically starting from the generalized Laplacian marginal prior. Significant advantages of this new approach are that it
does not depend on any preliminary edge detection steps, it is simpler to faster while it yields better results than the more complex ones. This is a new subband adaptive shrinkage function, which shrinks each coefficient according to the probability that it presents a signal of interest. Experimental results indicate that for natural images this estimator outperforms, in terms of MSE, any classical soft thresholding rule with a uniform threshold per subband. The results demonstrate that the new method outperforms other methods.

Here in this algorithm it is assumed that the input image is corrupted with additive white Gaussian noise of zero mean and variance. An orthogonal wavelet transformation [15] of the noisy input yields an equivalent additive white noise model in the transform domain. In each wavelet subband at a given scale and orientation, thus

\[ y_i = \beta_i + \epsilon_i, \ i = 1, 2, 3, \ldots, n \]  

Eq. 6.22

where \( \beta_i \) are noise-free wavelet coefficients, \( \epsilon_i \) are independently identically distributed (IID) normal random variables, which are statistically independent from \( \beta_i \) and \( n \) is the number of coefficients in a subband. For compactness, this algorithm suppressed here the indices that denote the scale and the orientation and denoted the spatial position with a single index, like in a raster scanning. We have suppress the spatial index too because the same shrinkage rule will be applied to all the coefficients in a given subband. Our approach is aimed for priors that are sharply peaked at zero and heavy tailed like Laplacian, generalized Laplacian and alpha-stable distributions. The generalized Laplacian also called generalized Gaussian prior for the noise-free subband data [4], [53] is

\[ f(\beta) = \frac{\lambda^\nu}{2\Gamma(\frac{\nu}{\lambda})} \exp\left(-\frac{\beta^\nu}{\lambda}\right) \]  

Eq. 6.23

where \( \Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \)

where is the Gamma function, \( \lambda > 0 \) is the scale parameter, and \( \nu \) is the shape parameter, which is typically \( \nu \in [0,1] \) for natural images.
Let us define a signal of interest as a noise-free coefficient component that exceeds a specific threshold and formulate the following two hypotheses: 

\[ H_0: |\beta| \leq T \]  

and  

\[ H_1: |\beta| > T \]  

Let \( P(H_1/y) \) denote the conditional probability that a wavelet coefficient contains a signal of interest, given its observed value. The Bayes rule yields

\[ P(H_1/y) = \frac{\mu \eta}{1 + \mu \eta} \]  

Eq. 6.26

where \( \mu = P(H_1)/P(H_0) \) and \( \eta = f(y/H_1)/f(y/H_0) \) and the product is called the generalized likelihood ratio. We now consider a simple shrinkage rule

\[ \beta = \frac{P(H_1/y)\mu \eta}{1 + \mu \eta} y \]  

Eq. 6.27

This is the equation for Probshrink threshold estimation.

The conditional density of noise free coefficients given \( H_0 \) (signal of interest is absent) is in other words the probability density function of insignificant noise free coefficients and that it is proportional to \( f(\beta) \) for \( \beta \leq T \) and equal to zero otherwise. Analogously, the conditional density of \( \beta \) given \( H_1 \) is the density of significant noise free coefficients and is proportional to \( f(\beta) \) for \( \beta > T \) and equal to zero otherwise. For the assumed additive noisemodel where the noise coefficients and the noise-free coefficients are, respectively, realizations of two stochastic processes that are statistically independent, the conditional densities of the noisy coefficients, \( f(y/H_0) \) and \( f(y/H_1) \) result from the following convolutions [52]

\[ f(y/H_0) = \int_{-\infty}^{\infty} \phi(y - \beta; \sigma) f(\beta H_0) d\beta \]  

Eq. 6.28

\[ f(y/H_1) = \int_{-\infty}^{\infty} \phi(y - \beta; \sigma) f(\beta H_1) d\beta \]  

Eq. 6.29
where $\phi(y; \sigma)$ is the zero mean Gaussian density with the standard deviation $\sigma$. It is observed that the smallest coefficients are heavily shrunk toward zero while the largest ones tend to remain unchanged. Between these extremes, there is a smooth transition which depends on the global subband statistics expressed through $P(H_1)$.

The first novelty of the proposed subband adaptive shrinkage method is the way this algorithm estimate the prior probability of signal presence. In related approaches it has been usually assumed $P(H_1)=P(H_0)=0.5$ was estimated empirically as a given fraction of the observed noisy coefficients. Here, this algorithm derive the probability $P(H_1)$ from the prior model for the noise-free coefficients in a given sub band. In particular, hypotheses model in equation (6.24) and (6.25) is the probability of signal presence amounts to the area under the tails of $f(\beta)$ for $\beta > T$ and thus, estimation of $p(H_1)$ as

$$P(H1) = 1 - \int_{-T}^{T} f(\beta) d\beta$$

Eq. 6.30

Next, step is to develop this expression for the generalized Laplacian prior and analyze the performance of the resulting ProbShrink rule in equation 6.27.

6.4.5.1 Estimation of ProbShrink for Generalized Laplacian Prior

Under the assumed prior in equation 6.23, the conditional densities of noise-free coefficients are

$$f(\beta H_0) = \begin{cases} B_0 \exp(-\lambda \beta^\nu), & \text{if } \beta \leq T \\ 0, & \text{if } \beta > T \end{cases}$$

Eq. 6.31

$$f(\beta H_1) = \begin{cases} 0, & \text{if } \beta \leq T \\ B_1 \exp(-\lambda \beta^\nu), & \text{if } \beta > T \end{cases}$$

Eq. 6.32

with the normalization constants

$$B_0 = \frac{\lambda \nu}{2 \Gamma(\frac{\nu}{\nu}) \Gamma inc((\lambda T)^\nu, \frac{1}{\nu})}$$

Eq. 6.33
\[ B_1 = \frac{\lambda \nu}{2\Gamma(\frac{1}{\nu})[1\Gamma_{inc}(\lambda T)^{\nu}]} \]

where \( \Gamma_{inc}(x; a) = (1/T(a)) \int_0^x t^{a-1} e^{-t} dt \) is the incomplete gamma function. From equation (6.29), thus

\[ P(H_1) = 1 - \Gamma_{inc}(\lambda T)^{\nu}, (1/\nu) \]

and, thus

\[ \mu = \frac{P(H_1)}{P(H_0)} = \frac{1-\Gamma_{inc}(\lambda T)^{\nu}^{\frac{1}{\nu}}}{\Gamma_{inc}(\lambda T)^{\nu}^{\frac{1}{\nu}}} \]

For the Laplacian prior (\( \nu = 1 \)), the above expression reduces to

\[ \mu = P(H_1)/P(H_0) = \exp(-\lambda T)/[1 - \exp(-\lambda T)] \]

Together with the likelihood ratio \( f(y H_1/f(y H_0) \), which is calculated using equation 6.30 and 6.31. This completes the specification of the sub band adaptive estimator for probShrink[53]

### 6.4.7 Estimation of the Laplacian Prior Parameters

In the proposed method, the parameters \( \lambda \) and \( \nu \) of the generalized Laplacian prior for noise-free data are estimated from the noisy histogram in each sub band. In particular, the variance \( \sigma_y^2 \) and the fourth moment \( m_{4,y} \) of the generalized Laplacian signal corrupted by additive white Gaussian noise with standard deviation \( \sigma \) are

\[ \sigma_y^2 = \sigma^2 + \frac{\Gamma(\frac{3}{\nu})}{\lambda^2 \Gamma(\frac{1}{\nu})} \quad m_{4,y} = 3\sigma^4 + \frac{6\sigma^2 \Gamma(\frac{3}{\nu})}{\lambda^2 \Gamma(\frac{1}{\nu})} + \frac{\Gamma(\frac{5}{\nu})}{\lambda^4 \Gamma(\frac{1}{\nu})} \]

From the above equation,

\[ k_y = \frac{\Gamma(\frac{1}{\nu}) \Gamma(\frac{2}{\nu})}{\Gamma(\frac{2}{\nu})} = \frac{m_{4,y} + 3\sigma^4 - 6\sigma^2 \sigma_y^2}{(\sigma_y^2 - \sigma^2)^2} \]

And

\[ \lambda = \left( \frac{\sigma_y^2 - \sigma^2}{\Gamma(\frac{1}{\nu}) \Gamma(\frac{2}{\nu})} \right) \]

\[ \text{Eq. 6.34} \]

\[ \text{Eq. 6.35} \]

\[ \text{Eq. 6.36} \]

\[ \text{Eq. 6.37} \]

\[ \text{Eq. 6.38} \]
The expression $\Gamma\left(\frac{1}{6}\right) \Gamma\left(\frac{5}{6}\right)/\Gamma^2\left(\frac{3}{6}\right)$ in the left hand side of equation 6.37 is a monotonic decreasing function of $\nu$. We solve the parameter $\nu$ numerically from this equation. Using the estimated value of the shape parameter, the scale parameter $\lambda$ follows directly from equation 6.37. In this method, Haar wavelets transform is used.

6.5 Flowchart of the Proposed Method
Figure 6.3 Flowchart of the Proposed Method

A

1. Estimation of Common Vector
2. Design of average vector from all common vectors.
3. Estimation of components in the direction of average common vector for all segments.
4. Calculate the Eigen vectors From Covariance matrix
5. Calculate the variance of the calculated components.
6. Calculate the optimum threshold by using ProbShrink rule of threshold estimation
7. Calculate PSNR, MSE and Execution
Flowchart of the proposed algorithm is shown in figure 6.3 presents the design and implementation of a PCA and Wavelet based image denoising technique. In this method, the input test image is subjected to additive white Gaussian noise having different noise variance levels. The noisy image is decomposed into different blocks. Then the calculation of principal components and their vectors for all the components is being done. As noise is having some common characteristics, the main task is search for the common and similar vectors in all the segments. [6] Then the next job is to design and develop average vector from all the common vector and estimate the components in the direction of average common vector for all the segments. After getting the common vector, covariance matrix is designed and the calculation of eigen vectors and Eigen values is done. Apply the PCA transform and remove the above calculated components from the respective segments and deriving the new dimensional data set[81][83]. The key idea of Principal Component Analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the Principal Components (PCs), which are uncorrelated, and which are ordered so that the first few components retain most of the variation present in all of the original variables. Then calculate the variance of the above calculated components and finding the optimum threshold by using different wavelet shrinkage rule of threshold estimation using wavelet domain filtration with variance estimated. Here this method have used ProbShrink rule for threshold estimation.