Image Denoising using Principal Component Analysis in Wavelet Domain and Total Variation Regularization in Spatial Domain

Brajesh Kumar Sahu
Department of Electronics & Communication
Samrat Ashok Technological Institute, Vidisha

Preety D. Swami
Department of Electronics & instrumentation
Samrat Ashok Technological Institute, Vidisha

ABSTRACT
This paper presents an efficient denoising technique for removal of noise from digital images by combining filtering in both the transform (wavelet) domain and the spatial domain. The noise under consideration is AWGN and is treated as a Gaussian random variable. In this work the Karhunen-Loeve transform (PCA) is applied in wavelet packet domain that spreads the signal energy in to a few principal components, whereas noise is spread over all the transformed coefficients. This permits the application of a suitable shrinkage function on these new coefficients and elimination of noise without blurring the edges. The denoised image obtained by using the above algorithm is processed again in spatial domain by using total variation regularization. This post processing results in further improvement of the denoised results. Experimental results show better performance in terms of PSNR as compared to the performance of the methods when incorporated individually.

Keywords
Image denoising, Principal component analysis, Total variation regularization, Wavelet packet transform.

1. INTRODUCTION
Real world images do not exist without the noise. Principal sources of noise in the digital images arise during image acquisition and transmission. The amount of noise depends on the different factors such as CCD cameras, light levels and sensor temperature. During transmission, images are corrupted mainly due to interference in the channel used for transmission. The mathematical model of noisy image is given by

$$\bar{f} = f + \eta$$

(1)

Where $\bar{f}$ is the noisy image, f is the original image and η is the additive white Gaussian noise with zero mean and standard deviation σ. All denoising methods are a type of low pass filtering. Filtering can be done in the spatial domain or in the transform domain. Filtering in the transform domain is more efficient and introduces fewer artifacts. The traditional linear filtering methods in spatial domain are Wiener filtering and mean filtering, whereas nonlinear filtering can be done through median filtering [1,2]. The drawback of these filtering methods is that they do not preserve the edges in the denoised image. Donoho and Johnstone introduced denoising via wavelet thresholding [3,4,5] and it is now widely applied in science and engineering applications. It is based on thresholding (adaptive or nonadaptive)

of the wavelet coefficients obtained from the orthogonal discrete wavelet transform (DWT) of data. The wavelet coefficients in the frequency detail subbands can be thresholded by means of soft/hard thresholding operator. Another image denoising scheme is by using Principal Component Analysis (PCA) [6,7]. It transforms the original data set in to PCA domain and by preserving only the most significant principal components, the noise and trivial information can be removed. In [8], PCA based method was proposed for image denoising using local pixel grouping. Drawback of this scheme is that due to its computational complexity it is very slow.

The proposed work combines the good properties of wavelet packet analysis with those of principal component analysis, realized using the Karhuen-loeve (KL) transform in wavelet domain. The KL transform in wavelet packet domain fully decorrelate the frequency subbands, wherein the energy of the signal is clustered in few principal components while the noise energy spread over all transformed coefficients. This allows to apply a suitable shrinkage function to these new coefficients, thus removing noise without blurring the edges. The resulting image is processed again in spatial domain by using total variation regularization.

Rest of the paper is organized as follows: Section 2 presents the general theory of wavelet packet decomposition. Section 3 briefly reviews the procedure of PCA. Image denoising using total variation regularization is presented in Section 4. Section 5 is devoted for the description of the proposed algorithm and Section 6 presents the experimental results and comparison. Finally, Section 7 concludes the paper.

2. WAVELET PACKET DECOMPOSITION
Wavelet packet decomposition (WPD) [9], also known as wavelet packets or subband tree is a wavelet transform where the discrete-time signal is passed through a larger number of filters than the discrete wavelet transform (DWT). In DWT, each level of decomposition is calculated by passing only the previous wavelet approximation coefficients (cA_j) through discrete-time low and high pass quadrature mirror filters. However, in WPD at any scale j both the approximation coefficients and the detail coefficients, {Horizontal(cH_j), Vertical(cV_j), and Diagonal(cD_j)} are decomposed to create a full binary tree. The wavelet packet decomposition at the first level ($j=1$) is shown in Fig.1. In the figure, f is the input signal (2D) and g, h are the low pass and high pass quadrature mirror filters respectively.
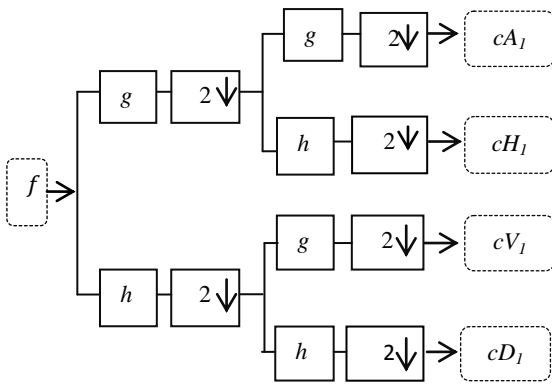
Fig 1: First level wavelet packet decomposition

3. PRINCIPAL COMPONENT ANALYSIS

PCA [6,10,11] is a classical decorrelation technique which is mostly used for data compression, patter recognition and noise reduction. A remarkable property possessed by PCA is that the energy of signal concentrates on a small subset of PCA transformed data set while the noise energy spreads over the whole dataset. Therefore, signal and noise can be distinguished in the PCA transform domain.

Assume $z = [z_1, z_2, z_3 \dots z_m]$ to be an m- component vector variable. Z, the $m \times n$ sample matrix of z, is given by

$$ Z \quad = \quad \begin{bmatrix} z_1^1 & z_1^2 & \dots & z_1^n \\ z_2^1 & z_2^2 & \dots & z_2^n \\ \vdots & \vdots & \vdots & \vdots \\ z_m^1 & z_m^2 & \dots & z_m^n \end{bmatrix} $$

(2)

For $i = 1, 2, 3 \dots m$ and $j = 1, 2, 3 \dots n$, z_i^j is the discrete sample of variable z_i. The ith row of sample matrix given by $Z_i = [z_i^1 \quad z_i^2 \dots z_i^n]$, is the sample vector of z_i. The mean value of z_i is estimated as $\mu_i = E[z_i] \approx \left(\frac{1}{n}\right)\sum_{j=1}^{n} Z_i(j)$. Thus the mean value vector of z is given by

$$ \mu = E[z] = [\mu_1 \ \mu_2 \dots \mu_m]^T $$

(3)

Centralized vector $\bar{z} = z - \mu$, and the element of $\bar{z}$ is $\bar{z}_i = z_i - \mu_i$. The sample vector of $\bar{z}_i$ is $\bar{Z}_i = Z_i - \mu_i = [\bar{z}_i^1 \bar{z}_i^2 \dots \bar{z}_i^n]$, where $\bar{z}_i^j = z_i^j - \mu_i$. Accordingly the centralized matrix $\bar{Z}$ of Z is given by

$$ \bar{Z} \quad = \begin{bmatrix} \bar{Z}_1 \\ \bar{Z}_2 \\ \vdots \\ \bar{Z}_m \end{bmatrix} = \begin{bmatrix} \bar{z}_1^1 & \bar{z}_1^2 & \dots & \bar{z}_1^n \\ \bar{z}_2^1 & \bar{z}_2^1 & \dots & \bar{z}_2^n \\ \vdots & \vdots & \vdots & \vdots \\ \bar{z}_m^1 & \bar{z}_m^2 & \dots & \bar{z}_m^n \end{bmatrix} $$

(4)

The co-variance matrix of $\bar{Z}$ is calculated as

$$ \Omega \qquad \approx \qquad (1/n)\bar{Z}\bar{Z}^T $$

(5)

In PCA transformation, an orthonormal transformation matrix P is calculated to decorrelate $\bar{Z}$, i.e. $\bar{X} = P\bar{Z}$ such that the co-variance matrix of $\bar{X}$ is diagonal. The co-variance matrix Ω of $\bar{Z}$ is symmetrical so it can be written as

$$ \Omega = \Phi\Lambda\Phi^T $$

(6)

where, $\Phi = [\phi_1\phi_2 \dots \dots \phi_m]$ is the $m \times m$ orthonormal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ is the diagonal eigenvalue matrix with $\lambda_1 \geq \lambda_2 \dots \dots \geq \lambda_m$. The terms $\phi_1, \phi_2, \dots \dots, \phi_m$ and $\lambda_1, \lambda_2, \dots \dots, \lambda_m$ are the eigenvectors and eigenvalues of Ω respectively. By putting

$$ P = \Phi^T $$

(7)

$\bar{Z}$ can be decorrelated, i.e. $\bar{X} = P\bar{Z}$ and $\Lambda = (1/n)\bar{X}\bar{X}^T$.

4. TOTAL VARIATION REGULARIZATION

Total variation, also known as the total variation regularization process is based on the principle that the signals with excessive and possible spurious details have high total variation, that is, integral of absolute gradient of the signal is high. By using this principle, upon reducing the total variation of the signal results in a close match to the original signal, thus removing undesired details (noise) while preserving the important details such as edges. For an input image $\hat{f}$ the goal of total variation is to estimate f'' that has a smaller total variation than $\hat{f}$ (noisy image) but is close to $\hat{f}$. One measure of closeness is sum of square error.
Total variation denoising problem is of the form proposed in [12].

$$ \min_{f''} \ \frac{l_m}{2} \left\| \hat{f} - f'' \right\|^2 + v(f'') $$

(8)

Where

$$ v(f'') = \sum_{i,j} \sqrt{\left| f''_{i+1,j} - f''_{i,j} \right|^2 + \left| f''_{i,j+1} - f''_{i,j} \right|^2} $$

(9)

and l_m is the regularization parameter, which plays an important role in denoising. If $l_m = 0$, there is no denoising. Depending upon the noise level, l_m can have a value between 0 and 1.

5. PROPOSED DENOISING ALGORITHM

In this section the proposed denoising algorithm is discussed. The original image f, of size $m \times m$ is corrupted by additive white Gaussian noise (η) with zero mean and standard deviation σ. The noisy image is represented by $\bar{f}(\bar{f} = f + \eta)$. Denoising is performed in two steps; Step 1: Wavelet packet denoising using PCA, Step 2: Post processing of the denoised image of step1 through the Total Variation Regularization in the spatial domain. The block diagram of the proposed technique is shown in Fig. 2.

Step 1: In the first step, the wavelet packet decomposition of noisy image is obtained by using the well-known tensor product technique. The l level decomposition of the noisy image ($\bar{f}$) of size $m \times m$ turns out in to $T = 2^l \times 2^l$ subband images represented by c_j ($j = 1, 2, \dots 2^l \times 2^l$), each of size of $k = m/2^l \times m/2^l$. Each sub image is arranged in the form of a single row of matrix Z. Thus, $2^l \times 2^l$ subbands when arranged in the matrix Z, makes the dimension of the matrix to be $(2^l \times 2^l) \times (m/2^l \times m/2^l)$. Then KL (PCA) transform is applied to the matrix Z and the transformed matrix is represented by X. The PCA transformed wavelet packet coefficients are represented by $\bar{c}_{j,k}$. The KL transform separates the signal and

noise energy, thus allowing to apply a new thresholding rule as proposed in [13]. All the frequency sub bands whose mean energy is less than the Gaussian noise variance (σ^2) are discarded and an exponential shrinkage function as proposed in [14] is applied to the retained coefficients.

$$F_t(x) = \begin{cases} a^{-h\frac{t^2}{x^2}} & \text{if } x \neq 0 \text{ with} \quad a \geq 1 \,\& h \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The threshold t is band dependent which means that each subband (c_j) has an individual threshold depending on the mean energy of subband (i.e., λ_j) and is given by t_j for j =1, 2, 3……. ($2^l \times 2^l$).

$$t_j = \sqrt{\frac{\lambda_j \sigma^2}{\lambda_j - \sigma^2}}$$

(11)

The thresholded coefficient is given by $\bar{\bar{c}}_{j,k}$ [$j = 1, 2, … (2^l \times 2l$ and $k=1, 2,…m/2l \times m/2l$.

$$\bar{\bar{c}}_{j,k} = \begin{cases} F_{t_j}(\bar{c}_{j,k}) \times \bar{c}_{j,k} & \text{if } \frac{1}{k}\bar{c}_j\bar{c}_j^T \geq \sigma^2 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

Application of inverse PCA and then the inverse wavelet packet transform to the thresholded coefficients results in the denoised image of step 1.

Step 2: In the second step, the resulting denoised image $\hat{f}$ of step 1 is treated as input image for denoising in the spatial domain by using the total variation regularization using Equation (8), and Chambolle's minimization algorithm proposed in [15].
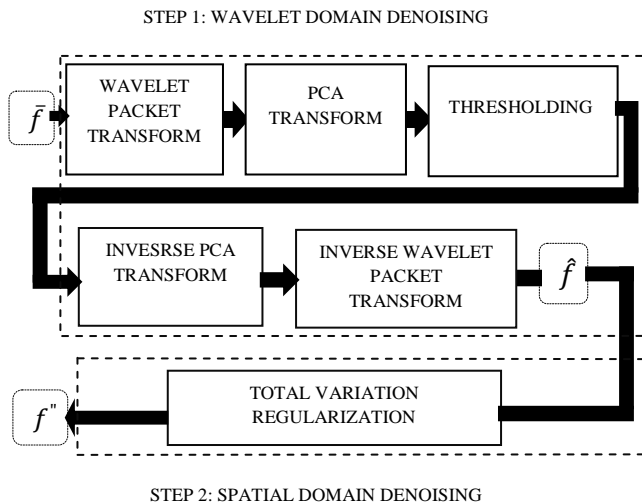
STEP 1: WAVELET DOMAIN DENOISING

STEP 2: SPATIAL DOMAIN DENOISING

Fig 2: Block diagram of proposed denoising algorithm.

6. EXPERIMENTAL RESULTS AND COMPARISON

The performance of the proposed denoising algorithm is tested on four gray-scale images shown in Fig. 3, each having a size of 512×512. These test images have been corrupted with Gaussian noise with different values of standard deviation ($\sigma = 5$, $\sigma = 10$, $\sigma = 15$, $\sigma = 25$, $\sigma = 35$). In the first step of the proposed

denoising algorithm, 3 level (*l=3*) wavelet packet decomposition is performed using 'sym8' wavelet for decomposition instead of 'db4' used by the authors in [13]. The near symmetry property of 'sym8' wavelet finds advantage over 'db4' in image processing applications as symmetry is a desirable property for the human visual system. For exponential shrinkage, a possible choice for the parameters *h* and *a* is *h=4* and *a=2* (generally these values depend on the amount of noise). In the second step of denoising, the denoised images obtained from the first step are further processed with TV regularization with differing values of regularization parameter l_m. Each image with standard deviation of noise (σ = 35, 25, 15, 10, 5) were tested for all possible values of l_m ranging from 0.1 to 0.9. Experiments show that, higher value of l_m is required for post-processing of images that were suffering from low noise levels and vice-versa. TV regularization parameter taken is l_m = 0.1, 0.2, 0.5, 0.7, 0.9 for the noise deviation σ = 35, 25, 15, 10, 5 respectively. Table 1 shows the experimental results in terms of root mean square error (RMSE) and peak signal to noise ratio ($PSNR = 20\log_{10}(255/RMSE)$). Table 2 shows the comparison of denoised results of proposed method with the soft thresholding (using the universal threshold) method, TV regularization method and PCA in wavelet domain method in terms of RMSE and PSNR. It is very clear from the comparative table that only for the low noise deviation ($\sigma = 5$) , PSNR of denoised Lena image of proposed method is low as compared to TV regularization method while for higher noise deviations, it is the highest as compared to all the methods under comparison. The PSNR of denoised cameraman and MRI image of proposed method for all noise deviations is high as compared to all the methods under comparison. For the denoised peppers image of proposed method, the PSNR is low as compared to TV regularization method for low noise deviation ($\sigma = 5$, $\sigma = 10$, $\sigma = 15$), but for higher noise deviations ($\sigma = 25$, $\sigma = 35$), the proposed method provide the best results. Denoised images of Lena, Cameraman, MRI, and Pepper by using different methods (TV regularization, PCA in wavelet domain, and proposed method) are shown in Fig. 4, Fig. 5, Fig. 6 and Fig. 7 respectively.
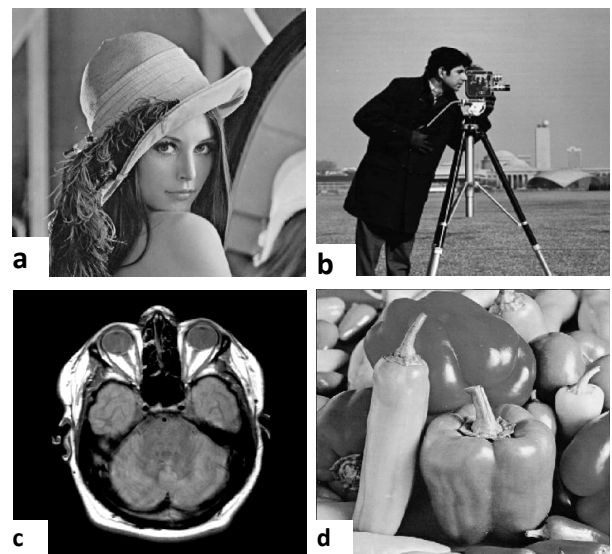

Fig3: Test images (a) Lena; (b) Cameraman; (c) MRI; (d) Peppers

Table 1. Numerical results on the test images with several noise standard deviation values.

Noise deviation (σ) and TV regularization parameter (l_m)		Lena		Cameraman		MRI		Peppers	
		RMSE	PSNR	RMSE	PSNR	RMSE	PSNR	RMSE	PSNR
σ = 5	l_m = 0.9	3.35	37.63	2.56	39.97	2.50	40.17	3.40	37.50
σ = 10	l_m = 0.7	4.99	34.16	4.11	35.86	3.82	36.49	5.00	34.15
σ = 15	l_m = 0.5	6.27	32.19	5.41	33.47	4.94	34.26	6.22	32.26
σ = 25	l_m = 0.2	8.22	29.83	7.60	30.52	6.87	31.39	7.90	30.18
σ = 35	l_m = 0.1	9.76	28.34	9.43	28.64	8.50	29.54	9.31	28.75

Table 2. Comparison between the soft thresholding with universal threshold method [4], TV regularization method [15], PCA in wavelet domain method [13], and proposed method in terms of RMSE and PSNR.

Methods	Soft thresholding with Universal threshold [4]		TV regularization Method [15] with regularization parameter (l_m)			PCA in wavelet domain method [13]		Proposed method	
	RMSE	PSNR	RMSE	PSNR	l_m	RMSE	PSNR	RMSE	PSNR
Lena									
σ = 5	4.44	35.18	3.14	**38.19**	0.4	3.67	36.83	3.35	37.63
σ = 10	6.45	31.94	5.11	33.96	0.1	5.41	33.46	4.99	**34.16**
σ = 15	8.85	29.46	6.10	32.42	0.1	6.70	31.61	6.27	**32.19**
σ = 25	10.43	27.76	11.88	26.63	0.1	8.78	29.26	8.22	**29.83**
σ = 35	12.20	26.40	20.19	22.03	0.1	10.44	27.75	9.76	**28.34**
Cameraman									
σ = 5	4.06	35.96	2.81	39.16	0.3	2.62	39.76	2.56	**39.97**
σ = 10	6.10	32.42	4.69	34.71	0.1	4.29	35.48	4.11	**35.86**
σ = 15	8.31	29.73	5.83	32.82	0.1	5.70	33.01	5.41	**33.47**
σ = 25	11.09	27.23	11.82	26.68	0.1	8.16	29.90	7.60	**30.52**
σ = 35	12.99	25.85	20.23	22.01	0.1	10.18	27.97	9.43	**28.64**
MRI									
σ = 5	3.76	36.64	2.54	40.04	0.3	2.69	39.52	2.50	**40.17**
σ = 10	5.84	32.80	4.15	35.76	0.1	4.29	35.46	3.82	**36.49**
σ = 15	8.10	29.96	5.64	33.10	0.1	5.56	33.22	4.94	**34.26**
σ = 25	9.94	28.18	11.87	26.64	0.1	7.71	30.85	6.87	**31.39**
σ = 35	11.86	26.65	20.30	21.98	0.1	9.50	28.57	8.50	**29.54**
Peppers									
σ = 5	4.42	35.22	3.03	**38.50**	0.3	3.51	37.23	37.23	37.50
σ = 10	6.47	31.91	4.60	**34.88**	0.1	5.31	33.63	33.63	34.15
σ = 15	8.62	29.42	5.77	**32.90**	0.1	6.65	31.68	31.68	32.26
σ = 25	10.33	27.85	11.66	26.80	0.1	8.75	29.29	29.29	**30.18**
σ = 35	12.22	26.39	20.20	22.02	0.1	10.50	27.70	27.70	**28.75**

Fig 4: (a) Lena noisy image with σ = 25; (b) denoised image with TV regularization algorithm, PSNR=26.63; (c) denoised image with PCA in wavelet domain, PSNR=29.26; (d) denoised image with proposed algorithm, PSNR=29.83.
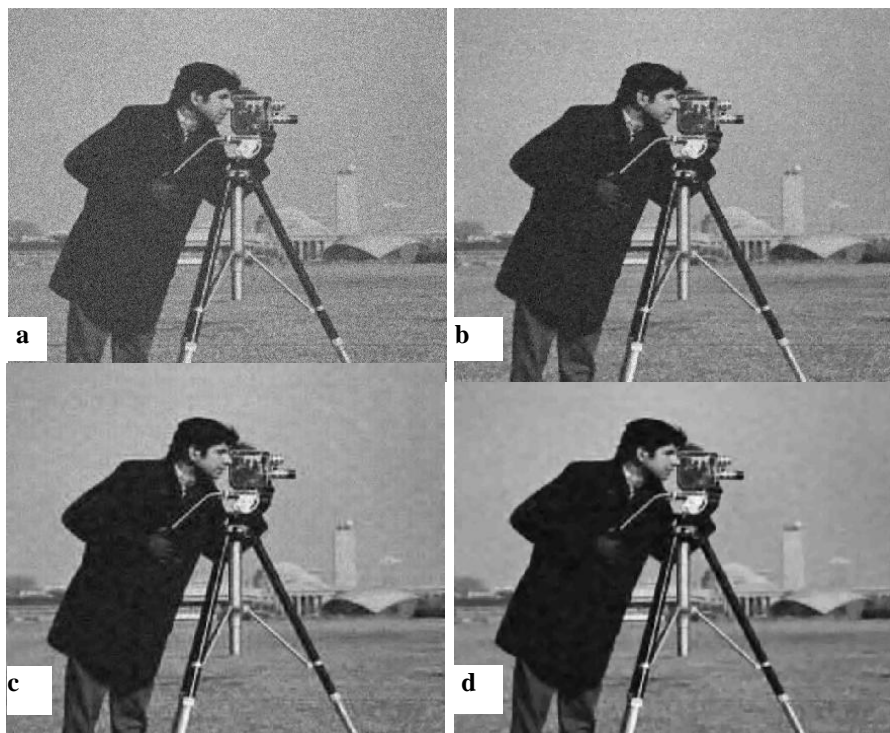

Fig 5: (a) Cameraman noisy image with σ = 25; (b) denoised image with TV regularization algorithm, PSNR=26.68; (c) denoised image with PCA in wavelet domain, PSNR=29.90; (d) denoised image with proposed algorithm, PSNR=30.52.
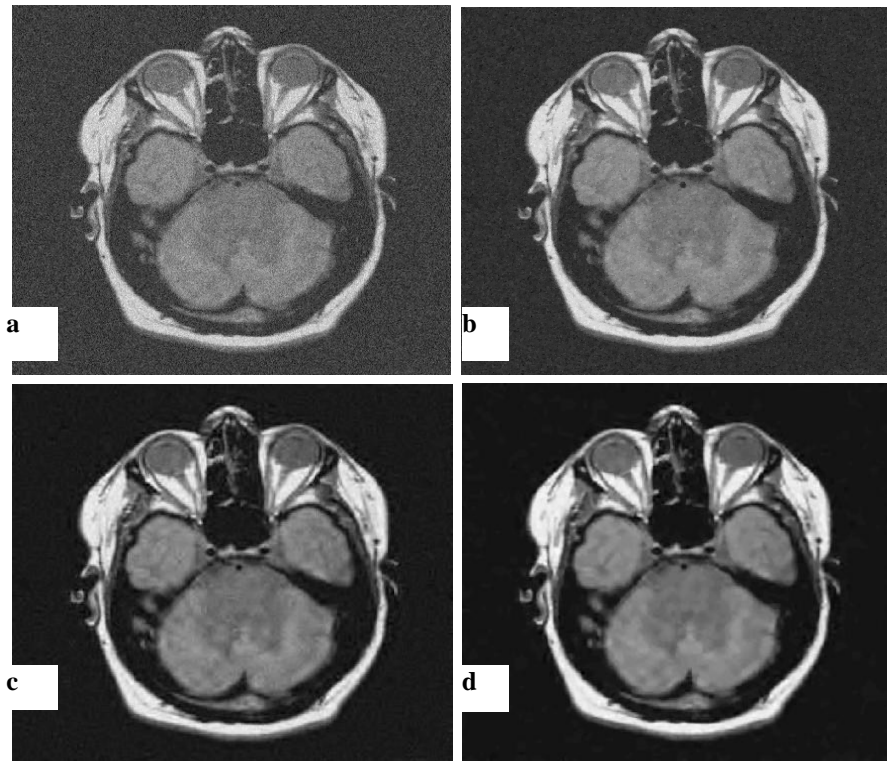
Fig 6: (a) MRI noisy image with σ = 25; (b) denoised image with TV regularization algorithm, PSNR=26.64; (c) denoised image with PCA in wavelet domain, PSNR=30.85; (d) denoised image with proposed algorithm, PSNR=31.39.
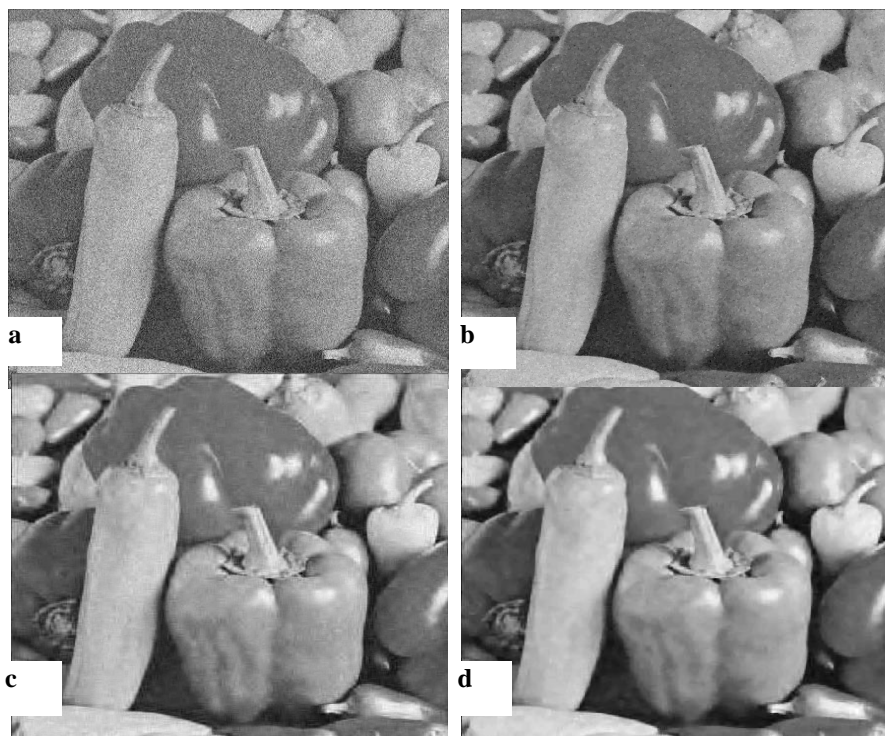

Fig 7: (a) Peppers noisy image with σ = 25; (b) denoised image with TV regularization algorithm, PSNR=26.80; (c) denoised image with PCA in wavelet domain, PSNR=29.29; (d) denoised image with proposed algorithm, PSNR=30.18.

7. CONCLUSION

This paper presents a denoising algorithm which is a combination of denoising in both the wavelet domain and the spatial domain. The first stage of denoising is performed by applying PCA on the wavelet packet transform coefficients. PCA owns the property of decorrelation and thus energy of signal concentrates on a small subset of PCA transformed data set while the noise energy spreads over the whole dataset. By applying a suitable shrinkage function, signal and noise are separated in the PCA domain. The second stage of denoising is a post processing step performed in spatial domain using total variation regularization. Experimental results demonstrated that the proposed method can effectively suppress the noise and at the same time effectively preserve the edges without blurring.

8. REFERENCES

[1] S. Jayaraman, S. Esakkirajan, and T. Veerakumar, *Digital image processing,* McGraw Hill, India, 2012.

[2] S. Dangeti, "Denoising technique comparison," M Tech thesis, Andhra University college of Engineering, Visakhapatnam, India, 2000.

[3] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via a wavelet shrinkage," *Biometrica,* vol. 81, pp. 425-455, 1994.

[4] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. on Information Theory,* Vol. 41, pp. 613-627, 1955.

[5] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via Wavelet shrinkage," *Journal of American Statistical Association,* Vol. 90, pp. 1200-1224, 1995.

[6] D. D. Muresan, T.W. Parks, "Adaptive principal components and image denoising," *International Conference on Image Processing,* Vol. 1, pp. 1101-1104, 2003.

[7] Y. Zu, Lecture notes on "Machine learning: Principal component analysis," 10-701, Spring 2012.

[8] L. Zhang, W. Dong, and D. Zhang, "Two stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition,* Vol. 43, pp. 1531-1549, 2010.

[9] R. C. Gonzalez, and R. E. Woods, *Digital image processing,* Pearson Education, india, 2010.

[10] A. Ghodsi, *Dimensionality reduction a short tutorial,* Department of statistics and actuarial science, University of Waterloo, Canada, 2006.

[11] A. Farag, and S. Elhabian, *A tutorial on principal component analysis,* University of Louisville, CVIP lab, 2009.

[12] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physican D,*Vol. 60, pp. 258-268, 1992.

[13] S. Bacchelli, and S. Papi, "Image denoising using principal component analysis in the wavelet domain," *Journal of Computational and Applied Mathematics,* Vol. 189, pp. 606-621, 2006.

[14] S. Bacchelli, and S. Papi, "Filtered wavelet thresholding methods," *Journal of Computational and Applied Mathematics,* Vol. 164-165, pp. 39-52, 2004.

[15] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision,* Vol. 20, pp. 89-97, 2004.