Non Local Means Image Denoising for Color Images Using PCA

P.A. Shyjila and M. Wilscy

Department of Computer Science, University of Kerala, Thiruvanathapuram
pashijila@gmail.com, wilsyphilipose@hotmail.com

Abstract. The goal of image denoising is to remove unwanted noise from an image. There are various methods for image denoising. The proposed algorithm is a variation of the nonlocal means (NLM) image denoising algorithm that uses principal component analysis (PCA) to achieve a higher accuracy while reducing computational load. Image neighborhood vectors are first projected onto a lower dimensional subspace using PCA. For color images RGB image neighborhood vectors are formed by concatenating image neighborhoods in the three color channels into a single vector. The dimensionality of this subspace is chosen automatically using parallel analysis. Consequently, neighborhood similarity weights for denoising are computed using distances in this subspace rather than the full space. The accuracy of NLM and the proposed algorithm are examined with respect to the choice of image neighborhood and search window sizes. Finally, we present a quantitative and qualitative comparison of the proposed algorithm versus NLM image denoising algorithm.

Keywords: Image denoising, nonlocal means (NLM), parallel analysis, principal component analysis, principal neighborhood.

1 Introduction

The goal of image denoising methods is to recover the original image from a noisy image. The best simple way to model the effect of noise on a digital image is to add a gaussian white noise. Several methods have been proposed to remove the noise and recover the true image. In most of them, denoising is achieved by averaging. This averaging may be performed locally (eg: the Gaussian smoothing model, the anisotropic filtering and the neighborhood filtering), by the calculus of variations (eg: the Total Variation minimization), or in the frequency domain (eg: the empirical Wiener filters and wavelet thresholding methods).

The denoising methods should not alter the original image . Now, most denoising methods degrade or remove the fine details and texture of the original method. Many models used in denoising applications have been based on the assumption of piecewise smoothness [2],[3] and [4]. This type of model is too simple to capture the textures present in a large percentage of real images. This drawback has limited the performance of such models, and motivated data driven representations. One data-driven strategy is to use image neighborhoods or patches as a feature vector for representing local structure. Image neighborhoods are rich enough to capture the local

N. Meghanathan et al. (Eds.): CCSIT 2011, Part I, CCIS 131, pp. 288–297, 2011.

structures of real images. This representation has been used as a basis for image denoising , for texture synthesis , and for texture segmentation. The image neighborhood feature vector is typically high dimensional. For instance, it is 49 dimensional if 7×7 neighborhoods are used. Hence, the computation of similarities between feature vectors incurs a large computational cost. By projecting image neighborhood vectors into a lower dimensional subspace using Principal Component Analysis(PCA) we can reduce the computational complexity of methods that rely on image neighborhood information.

Principal Neighborhood Dictionaries for Non Local Means imade denoising algorithm (PND) [1] is a variation of the nonlocal means (NLM) image denoising algorithm that uses principal component analysis (PCA) to achieve a higher accuracy while reducing computational load. We extend this method by applying it for color images .The nonlocal means (NLM) image denoising algorithm averages pixel intensities using a weighting scheme based on the similarity of image neighborhoods [6]. The nonlocal means (NLM) image denoising algorithm using PCA is a very similar approach in which image neighborhood vectors are projected to a lower dimensional subspace using principal component analysis (PCA). For color images, RGB image neighborhood vectors are formed by concatenating image neighborhoods in the three color channels into a single vector. Then, the neighborhood similarity weights for denoising are computed from distances in this subspace resulting in significant computational savings. More importantly, this approach results in increased accuracy over using the fulldimensional ambient space [9], [10].

One disadvantage of the approach in [10] is the introduction of a new free parameter to the algorithm—the dimensionality of the PCA subspace. This approach proposes an automatic dimensionality selection criteria using parallel analysis [15] that eliminates this free parameter. The parallel analysis method compares the eigen values of the data covariance matrix to eigen values of the covariance matrix of an artificial data set. The method for creating artificial data set is described in section 3. The detailed discussion of the related works are provided in the next section. We provide a detailed discussion of the proposed method in section 3. Finally in section 4 we provide the results and analysis. In this section we compare the proposed method and the NLM algorithm for color images.

2 Related Work

Buades *et al.* introduced the NLM image denoising algorithm which averages pixel intensities weighted by the similarity of image neighborhoods [6]. Image neighborhoods are typically defined as 5 ×5, 7× 7, or 9× 9 square patches of pixels which can be seen as 25, 49 or 81 dimensional feature vectors, respectively. Then, the similarity of any two image neighborhoods is computed using an isotropic Gaussian kernel in this high-dimensional space. Finally, intensities of pixels in a search-window centered around each pixel in the image are averaged using these neighborhood similarities as the weighting function.

Mahmoudi and Sapiro have proposed a method to improve the computational efficiency of the NLM algorithm [28]. Their patch selection method removes unrelated neighborhoods from the search-window using responses to a small set of predetermined filters such as local averages of gray value and gradients. Unlike [28] the lower dimensional vectors computed in [9], [10] and [11] are data-driven. Additionally, in [9] and [10], the lower dimensional vectors are used for distance computation rather than patch selection.

In the above methods, the computational complexity is very high because of high dimensionality image neighborhoods. In Principal Neighborhood Dictionaries for Non Local Means Image Denoising Algorithm (PND) [1], this computational complexity is reduced by projecting image neighborhood vectors into a lower dimensional subspace using Principal Component Analysis (PCA). We extend this method by applying it for color images.

Principal component analysis of neighborhoods have previously been used for various image processing tasks. PCA of image neighborhoods was used for denoising [5]. However, in that work, PCA is computed for local collections of image neighborhood samples and denoising is achieved by direct modification of the projection coefficients.

In this algorithm, PCA is computed once, globally rather than locally. This results in a computationally more efficient algorithm. Furthermore, a nonlocal means averaging scheme is used rather than direct modification of projection coefficients.

There are various methods for determining the number of components to retain in data analysis. Parallel Analysis, originally proposed by Horn [15], is one of the most successful methods for determining the number of true principal components. Improvements to the original parallel analysis method have also been proposed. In this algorithm, the method used in [1] is used for subspace dimensionality selection.

3 Proposed Method

The proposed method is a variation of the nonlocal means (NLM) image denoising algorithm. The nonlocal means (NLM) image denoising algorithm averages pixel intensities using a weighting scheme based on the similarity of image neighborhoods. The computational complexity of NLM is very high because of high dimensionality image neighborhoods. The proposed method and PND [1] are very similar approaches in which image neighborhood vectors are projected to a lower dimensional subspace using principal component analysis (PCA). For color images, RGB image neighborhood vectors are formed by concatenating image neighborhoods in the three color channels into a single vector. Then, the neighborhood similarity weights for denoising are computed from distances in this subspace resulting in significant computational savings. More importantly, this approach results in increased accuracy over using the full-dimensional ambient space [9], [10].

The nonlocal means (NLM) image denoising algorithm is discussed in section 3.1. In section 3.2, the proposed method is given. The method for determining the subspace dimensionality is given in section 3.3. Finally, section 3.4 describes how we select the size of search window.

3.1 Non Local Means Algorithm

Starting from a discrete image u, a noisy observation of u at pixel i is defined as $v(i) = u(i) + n(i)$. Let N_i denote a r × r square neighborhood centered around pixel

i. Also, let y(i) denote the vector whose elements are the gray level values of v at pixels in N_i. Finally, S_i is a square search-window centered around pixel i. Then, the NLM algorithm [6] defines an estimator for u(i) as

$$\mathrm{NL(i)} = \sum_{j\in S_i} \frac{1}{Z(i)} e^{-\|y(i)-y(j)\|^2/h^2} v(j) \tag{1}$$

where $Z(i) = \sum_{j\in S_i} e^{-\frac{\|y(i)-y(j)\|^2}{h^2}}$ is a normalizing term. The smoothing kernel width parameter h controls the extent of averaging. For true nonlocal means, the search window S_i needs to be the entire image for all i , which would give rise to global weighted averaging. However, for computational feasibility, S_i has traditionally been limited to a square window of modest size centered around pixel i. This is the limited-range implementation of the NLM algorithm as proposed in the pioneering work by [6]. For instance, a 21 × 21 window is used in [6] whereas a 7 × 7 window is used in [9].

3.2 Nonlocal Means Image Denoising for Color Images Using PCA

For color images, y(i)'s (image neighborhood vectors) are formed by concatenating image neighborhoods in the three color channels into a single vector. Also in this approach, the distances $\|y(i)-y(j)\|^2$ in equation (1) are replaced by distances computed from projections of y onto a lower dimensional subspace determined by PCA. Let Ω denote the entire set of pixels in the image. Also, let Ψ be a randomly chosen subset of Ω. Treating y(i) as observations drawn from a multivariate random process, we can estimate their covariance matrix as

$$C_y = \frac{1}{\psi}(y(i)-\bar{y})(y(i)-\bar{y})^T \tag{2}$$

where $\bar{y} = \frac{1}{|\psi|}\sum_{i\in\psi} y(i)$ is the sample mean and $|\psi|$ is the number of elements in the set ψ . A small subset $\psi \subseteq \Omega$ is typically sufficient to accurately estimate the covariance matrix and results in computational savings. The dimensionality of a r×r neighborhood vector is r^2 . For simplicity of notation, let $M= r^2$. Then C_y is a M×M matrix. Let $\{b_p : p = 1 : M\}$ be the eigenvectors of C_y , i.e., the principal neighborhoods, sorted in order of descending eigen values. Let the d-dimensional PCA subspace be the space spanned by $\{b_p : p = 1 : d\}$. Then the projections of the image neighborhood vectors onto this subspace is given by

$$y_d(i) = \sum_{p=1}^{d} \langle y(i), b_p \rangle b_p \tag{3}$$

where $\langle y(i), b_p \rangle$ denotes the inner product of the two vectors.

Let $f_d(i) = [\langle y(i), b_1 \rangle \ldots \langle y(i), b_d \rangle]^T$ be the d-dimensional vector of projection coefficients. Then, due to the orthonormality of the basis functions

$$\|y_d(i) - y_d(j)\|^2 = \|f_d(i) - f_d(j)\|^2 \tag{4}$$

Finally, define a new family of estimators for d∈[1,M]

$$\hat{u}_d(i,1) = \sum_{j \in S_i} \frac{1}{Z_d(i)} e^{-\frac{\|f_d(i) - f_d(j)\|^2}{h^2}} v(j,1) \tag{5}$$

$$\hat{u}_d(i,2) = \sum_{j \in S_i} \frac{1}{Z_d(i)} e^{-\frac{\|f_d(i) - f_d(j)\|^2}{h^2}} v(j,2) \tag{6}$$

$$\hat{u}_d(i,3) = \sum_{j \in S_i} \frac{1}{Z_d(i)} e^{-\frac{\|f_d(i) - f_d(j)\|^2}{h^2}} v(j,3) \tag{7}$$

where $Z_d(i) = \sum_{j \in S_i} e^{-\frac{\|f_d(i) - f_d(j)\|^2}{h^2}}$ is the new normalizing term.

3.3 Automatic Subspace Dimensionality Selection

The original parallel analysis method [15] compares the eigen values of the data covariance matrix to eigen values of the covariance matrix of an artificial data set. This artificial data set is generated by drawing samples from a multivariate normal distribution with the same dimensionality M, the same number of observations $|\psi|$, and the same marginal standard deviations as the actual data. Let λ_p for $1 \leq p \leq M$ denote the eigen values of C_y sorted in descending order. Similarly, let α_p denote the sorted eigen values of the artificial data covariance matrix. Parallel analysis estimates data dimensionality as

$$d = \max(\{1 \leq p \leq M \lambda_p \geq \alpha_p\}) \tag{8}$$

The intuition is that the P_3 is a threshold for λ_p below which the p^{th} component is judged to have occurred due to chance.

An improvement to parallel analysis is to use Monte Carlo simulations to generate the artificial data which removes the assumption of normal distribution. This algorithm generates the artificial data by randomly permuting each element of the neighborhood vector across the sample Ψ. Let $y_{i,k}$ denote the k^{th} element of the neighborhood vector y(i). For each k generate a random permutation j(i) of the sequence i= 1 : |Ψ| and let

$w_{i,k} = y_{j(i),k}$. Then, the random vectors w(i) are composed from the elements $w_{i,k}$. The artificial eigen values α are computed from the covariance matrix of w. This method for computing the artificial covariance matrix keeps the marginal distributions intact while breaking any interdependencies between them.

Several researchers have previously discussed that parallel analysis has a strong tendency to underestimate the number of components in data where the first component is much more significant than the rest of the components (oblique structure). This is the case with image neighborhoods where the first component, which is always approximately the average intensity in the neighborhood, has a much larger eigen value than the rest of the components. Therefore, this algorithm removes the effect of the first component. Here we compute the average intensity of the neighborhood $\mu_i = \frac{1}{M} \sum_{k \in N_i} y_{i,k}$ and generate a new set of neighborhood vectors whose elements are $y'_{i,k} = y_{i,k} - \mu_i$. Finally, the artificial data are generated from the permutations $y_{i,k} = y'_{j(i),k}$.

3.4 Smoothing Kernel Width Selection

The optimal choice of the parameter h in equations (5), (6) and (7) varies significantly with the image neighborhood size and choice of subspace dimensionality. We empirically find the optimal h for each combination of d and N for the set of test images used in this paper. This is repeated at various noise standard deviations added to the images. The optimal value of h behaves in a very predictable manner as a function of the noise level and PCA subspace dimensionality .

4 Results and Analysis

In this section, we present detailed experimental results studying the behavior of the proposed algorithm with respect to subspace dimensionality and search-window size selections. We also present quantitative and qualitative comparisons with the original NLM algorithm applied for color images. We study the performance of the proposed approach using images corrupted with additive, independent Gaussian noise with zero mean and standard deviations 10, 15 and 20.

In section 4.1. we present the comparison across various image neighborhood sizes. In section 4.2, we will compare the performance of the full proposed algorithm to the NLM algorithm for color images.

4.1 Subspace Dimensionality and Image Neighborhood Size

Comparison across various image neighborhood sizes are given in Table 1. For each test image, Table 1 includes three rows, one for each input noise level. Each row gives the best PSNR values at the optimal choice of d for different neighborhood sizes. Results for image neighborhoods ranging from 5×5 to 9×9 are provided.. Finally, the overall best PSNR across the various neighborhood sizes for a particular image and noise level is shown in boldface.

Table 1. Psnr values at the optimal subspace dimensionality for three noise levels (σ= 10, 15, 20) shown in column 3. Columns 4-6 show the results at the best for neighborhood sizes from 5×5 to 9×9.

Input Image	Noise Variances	PSNR of the noisy image	5×5 NLMP CA	7×7 NLMP CA	9×9 NLMP CA
Lena	10	32.05	34.40	**34.45**	34.45
	15	30.17	**33.04**	33.04	33.03
	20	29.30	**32.13**	32.03	32.13
Baloon	10	32.05	33.47	33.49	**33.52**
	15	30.21	32.18	**32.19**	32.17
	20	29.33	31.32	31.45	**31.47**
Island	10	32.13	**32.81**	32.80	32.79
	15	30.27	**31.34**	31.32	31.32
	20	29.38	30.65	**30.66**	30.61
House	10	32.05	33.81	33.83	**33.87**
	15	30.18	32.59	32.59	**32.61**
	20	29.29	31.81	**31.85**	31.85
Tree	10	32.04	32.34	**32.38**	32.25
	15	30.19	31.09	31.12	**31.18**
	20	29.29	**30.37**	30.35	30.37
Girl	10	32.01	35.31	35.31	**35.27**
	15	30.17	33.80	33.80	**33.80**
	20	29.31	33.00	33.00	**33.00**

As the noise level increases, weight reliability becomes increasingly important; hence, larger image neighborhoods are preferred.

Table 2. PCA subspace dimensionality selected by parallel analysis

Input Image	Noise Variances	5×5 NLMP CA	7×7 NLMP CA	9×9 NLMP CA
Lena	10	3	4	6
	15	3	4	6
	20	3	4	6
Baloon	10	3	4	5
	15	3	4	5
	20	3	4	5
Island	10	2	4	5
	15	2	4	5
	20	2	4	5
House	10	2	4	5
	15	2	4	4
	20	2	4	4
Tree	10	4	6	8
	15	4	6	8
	20	4	6	8
Girl	10	3	6	8
	15	3	6	8
	20	3	6	7

4.2 Comparison with NLM Algorithm for Color Images

Table 2 shows the values selected by parallel analysis for various neighborhood sizes and noise levels. Table 3 compares the results of the proposed algorithm with these automatically chosen and values to the results of the NLM algorithm for color images. Table 4 shows the time taken for denoising using the proposed algorithm and NLM.

Table 3. Psnr values for images denoised with the proposed algorithm and NLM for color images

Input Image	Noise Variances	5×5		7×7		9×9	
		NLMPCA	NLM	NLMPCA	NLM	NLMPCA	NLM
Lena	10	34.38	33.62	**34.42**	33.61	34.42	33.62
	15	**32.31**	31.42	32.31	31.42	32.31	31.42
	20	**30.96**	30.21	30.92	30.20	30.92	30.20
Baloon	10	33.45	33.10	33.18	33.09	**33.52**	33.09
	15	31.62	31.17	31.42	31.16	**31.63**	31.16
	20	**30.68**	30.08	30.14	30.08	30.62	30.08
Island	10	**32.82**	32.55	32.82	32.56	32.78	32.56
	15	31.26	30.99	**31.47**	30.98	31.44	30.99
	20	30.38	30.07	**30.57**	30.07	30.42	30.07
House	10	33.83	33.49	33.83	33.50	**33.85**	33.50
	15	31.78	31.39	31.81	31.38	**32.02**	31.39
	20	30.46	30.22	**30.79**	30.21	30.73	30.20
Tree	10	32.34	**32.45**	32.38	32.45	32.25	32.45
	15	31.09	30.96	31.12	30.96	**31.18**	30.95
	20	30.32	30.00	**30.35**	30.00	30.33	30.00
Girl	10	34.75	33.88	**34.78**	33.89	34.41	33.87
	15	31.84	31.54	32.20	31.55	**31.91**	31.53
	20	30.58	30.22	30.97	30.22	**30.98**	30.22

Table 4. Time for denoising using the proposed algorithm and NLM for color images of noise variance 10. Image neighborhood size is 7×7.

Input Image	Size	Time in seconds	
		NLMPCA	NLM
Lena	256×256×3	42.765	76.328
Baloon	171×256×3	28.266	51.329
Island	192×256×3	31.594	57.375
House	256×256×3	41.984	76.735
Tree	256×256×3	43.391	76.578
Girl	256×256×3	43.454	76.781

5 Conclusion

The proposed algorithm is a variation of the nonlocal means (NLM) image denoising algorithm that uses principal component analysis (PCA) to achieve a higher accuracy while reducing computational load. The accuracy and computational cost of the NLM image denoising algorithm is improved by computing neighborhood similarities, i.e., averaging weights, after a PCA projection to a lower dimensional subspace. We showed that parallel analysis can be used to automatically determine a subspace dimensionality that yields good results. Consequently, neighborhood similarity weights for denoising are computed using distances in this subspace rather than the full space. Then accuracy of NLM and the proposed algorithm are examined with respect to the choice of image neighborhood and search window sizes. Also the proposed algorithm is compared with NLM image denoising algorithm. This approach can also be easily applied to other denoising and segmentation algorithms that use similarity measures based on image neighborhood vectors.

References

[1] Tasdizen, T.: Senior Member, IEEE, Principal Neighborhood Dictionaries for Nonlocal Means Image Denoising. IEEE Trans. Image processing 18(12) (December2009)
[2] Mumford, D., Shah, J.: "Optimal approximations by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math. 42(4), 577–685 (1989)
[3] Perona, P., Malik, J.: "Scale-space edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. 12(7), 629–639 (1990)
[4] Rudin, L.I., Osher, S.J., Fatemi, E.: "Nonlinear total variation based noise removal algorithms. Phys. D 60, 259–268 (1992)
[5] Muresan, D.D., Parks, T.W.: Adaptive principal components and image denoising. In: Proc. Int. Conf. Image Processing, vol. 1, pp. 101–104 (2003)
[6] Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 60–65 (2005)
[7] Kervrann, C., Boulanger, J.: Optimal spatial adaptation for patchbased image denoising. IEEE Trans. Image Process. 15(10), 2866–2878 (2006)
[8] Awate, S.P., Whitaker, R.T.: Unsupervised, information-theoretic, adaptive image filtering for image restoration. IEEE Trans. Pattern Anal. Mach. Intell. 28(3), 364–376 (2006)
[9] Azzabou, N., Paragios, N., Guichard, F.: Image denoising based on adapted dictionary computation. In: Proc. Int. Conf. Image Processing, vol. 3, pp. 109–112 (2007)
[10] Tasdizen, T.: Principal components for non-local means image denoising. In: Proc. Int. Conf. Image Processing (2008)
[11] Orchard, J., Ebrahim, M., Wang, A.: Efficient nonlocal-means denoising using the SVD. In: Proc. Int. Conf. Image Processing (2008)
[12] Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: Proc. IEEE Int. Conf. Computer Vision, pp. 1033–1038 (1999)
[13] Wei, L.-Y., Levoy, M.: Fast texture synthesis using tree-structured vector quantization. In: Proc. SIGGRAPH, pp. 479–488 (2000)
[14] Awate, S.P., Tasdizen, T., Whitaker, R.T.: Unsupervised texture segmentation with nonparametric neighborhood statistics. In: Proc. Eur. Conf. Computer Vision, pp. 494–507 (2006)

[15] Horn, J.L.: A rationale and test for the number of factors in factor analysis. Psychometrica 30(2), 179–185 (1965)
[16] Weickert, J.: Anisotropic Diffusion in Image Proc. Verlag, Teubner (1998)
[17] Chan, T., Shen, J., Vese, L.: "Variational Pde models in image proc. Notice Amer. Math. Soc. 50, 14–26 (2003)
[18] Nordstrom, K.N.: Biased anisotropic diffusion: A unified regularization and diffusion approach to edge detection. Image Vis. Comput. 8(4), 318–327 (1990)
[19] Kindermann, S., Osher, S., Jones, P.W.: Deblurring and denoising of images by nonlocal functionals. Multiscale Model. Simul. 4(4), 1091–1115 (2005)
[20] Brox, T., Kleinschimdt, O., Cremers, D.: Efficient nonlocal means for denoising of textural patterns. IEEE Trans. Image Processing 17(7), 1083–1092 (2008)
[21] Glorfeld, L.W.: An improvement on horn's parallel analysis methodology for selecting the correct number of factor's to retain. Educ. Psych. Meas. 55(3), 377–393 (1995)
[22] Starck, J., Candes, E., Donoho, D.: The curvelet transform for image denoising. IEEE Trans. Image Process. 11(6) (June 2000)
[23] Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M.: A joint inter and intrascale statistical model for bayesian wavelet based image denoising. IEEE Trans. Image Process 11, 545–557 (2002)
[24] Sendur, L., Selesnick, I.: Bivariate shrinkage functions for waveletbased denoising exploiting interscale dependency. IEEE Trans. Signal Process. 50, 2744–2756 (2002)
[25] Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.: Image denoising using scale mixtures of gaussians in the wavelet domain. IEEE Trans. Image Process. 12, 1338–1351 (2003)
[26] Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionary. IEEE Trans. Image Process. 15(12), 3736–3745 (2006)
[27] Roth, S., Black, M.J.: Fields of experts: A framework for learning image priors with applications. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 860–867 (2005)
[28] Mahmoudi, M., Sapiro, G.: Fast image and video denoising via nonlocal means of similar neighborhoods. IEEE Signal Proc. Lett. 12(12), 839–842 (2005)