

ADAPTIVE PRINCIPAL COMPONENTS AND IMAGE DENOISING

D. Darian Muresan

Digital Multi-Media Design (DMMD)
1633 N. Quinn St., Suite 304
Arlington, VA. 22209
darian@dmmd.net

Thomas W. Parks

Electrical and Computer Engineering,
Cornell University
Ithaca, NY 14853
parks@ece.cornell.edu

ABSTRACT

This paper presents a novel approach to image denoising using adaptive principal components. Our assumptions are that the image is corrupted by additive white Gaussian noise. The new denoising technique performs well in terms of image visual fidelity, and in terms of PSNR values, the new technique compares very well against some of the most recently published denoising algorithms.

1. INTRODUCTION

This paper investigates the problem of image denoising when the image is corrupted by additive white Gaussian noise, which is a valid assumption for images obtained through scanning or other image capturing devices. A lot of work on noise reduction is based on wavelet thresholding [1], a simple and very effective denoising method. The basic idea is to project the noisy signal onto a properly selected orthogonal set of basis functions, such that the high frequency coefficients are mostly due to noise. Then, the small high frequency coefficients can be safely set to zero, preserving the structure of the original signal, while removing noise. Finding the best signal representation and the proper threshold is discussed in detail in [2].

A large percentage of the image denoising algorithms assume an orthogonal basis decomposition of the signal. While this may be an efficient way to decompose the image for compression purposes, several authors [3, 4, 5] have shown that an over-complete representation of the signal is superior for image denoising. The main advantage of over-complete expansion is summarized by [3] as a suppression of the Gibbs phenomena. In [3] the Translation-Invariant denoising algorithm is achieved by shifting the signal multiple times, denoising each shifted signal separately (using orthogonal decompositions for each shift), shifting back and then averaging the results. When denoising shifted versions of the signal, edge artifacts occur at different locations. When the signals are shifted back and

averaged these edge artifacts are averaged as well. The authors of [3] showed that a uniform thresholding in a Translation Invariant denoising does well in eliminating some of the edge artifacts seen in orthogonal wavelet denoising.

The authors in [5] extend the idea of [3] by simultaneously processing all the shifted versions to obtain more accurate statistical models for signal components. The work of [4] extends the idea of wavelet thresholding to an adaptive wavelet thresholding method based on context modeling. Each wavelet coefficient is modeled as a random variable of a generalized Gaussian distribution with an unknown parameter. Experimentally, their adaptive thresholding using shift-invariant non-subsampled wavelet transform (SI-AdaptShrink) is one of the best denoising algorithms.

All denoising algorithms reviewed are some form of a low pass filter. The assumption is that noise is captured by the high frequency coefficients and by filtering these coefficients the unwanted noise is removed. Unfortunately, edges also have high frequency components and by removing noise, high frequency components belonging to edges are also removed. This is accentuated when using separable wavelets, as is the case with most denoising algorithms in literature. By generating 2-D basis sets, which have vectors lined up along edges, and not across them, the high frequency coefficients caused by edges are much smaller. This in turn improves the denoising algorithm. The selection of 2-D locally adaptive basis sets is the main contribution of this paper.

2. THE ALGORITHM

Signal decompositions based on edge direction [6] decompose an image, based on both scale and local edge direction, using steerable filters. Our approach uses principal components (PC) on local image patches to derive a 2-D, locally adaptive basis set. The local principal components provide the best local¹ basis set and the largest eigenvector is in the

This work was supported by NSF MIP9705349, TI and Kodak

¹The basis set minimizes the sum of the squares of the errors between the first N basis and a set of training vectors, \mathcal{S} , representative of the local

direction of the local image edge. As it will be shown, this new basis decomposition is more efficient at reducing white Gaussian noise than using a wavelet decomposition. This is especially true for highly structured regions, such as the stripes on *barb*, where the local edge direction can be estimated much better. The idea behind Gaussian image denoising is to decompose the local signal using the locally adapted principal components (PC), threshold the coefficients, and then reconstruct. Our approach is similar to [7], except that we use principal components instead of independent components and the local basis functions are determined adaptively from the local image patches as opposed to being fixed for the entire image and then being applied adaptively to local image patches, as in [7].

Our image model is based on adaptively determining the best local basis and this means that PCs are recalculated at different image locations. To denoise an entire image all the patches that make up the image are denoised separately. The noisy image is given by

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$ is zero mean, white Gaussian noise with unknown variance σ^2 . The locally adaptive basis functions are the principal components of S , where S contains a local collection of noisy image blocks, \mathbf{y}_i , with $1 \leq i \leq M$. Noise removal is done by decomposing the signal into the local principal components, estimating the clean coefficients, and reconstructing. The approach to estimating the clean transform coefficients from the noisy ones follows the approach of [8] for estimating the clean wavelet coefficients from the noisy ones. In particular, the clean PC transform coefficients are obtained using an optimal LMMSE.

The noise variance σ^2 is estimated, as in [9], using the robust median estimator of the highest sub-band of a Daubechies Two wavelet transform:

$$\hat{\sigma} = \frac{\text{Median}(HH)}{0.6745}, \quad (2)$$

where HH are the high-high band wavelet coefficients. This approach of estimating σ^2 is also used in [4]. An alternative approach is to let $\sigma^2 = \text{Variance}(HH)$. Even better, is to use the local basis decomposition and look at the transform coefficients in the direction of the smallest principal components of S . Next, let x_i^l be the transform coefficient of vector \mathbf{x}_i in the direction of the l^{th} principal components of S (i.e. the projection of \mathbf{x}_i onto the l^{th} principal component), and similarly for y_i^l with vector \mathbf{y}_i , and n_i^l with vector \mathbf{n}_i . Then

$$y_i^l = x_i^l + n_i^l, \quad (3)$$

where $n_i^l \sim \mathcal{N}(0, \sigma^2)$. Recall that since the noise samples are independent and identically distributed (i.i.d.) Gaussian,

image.

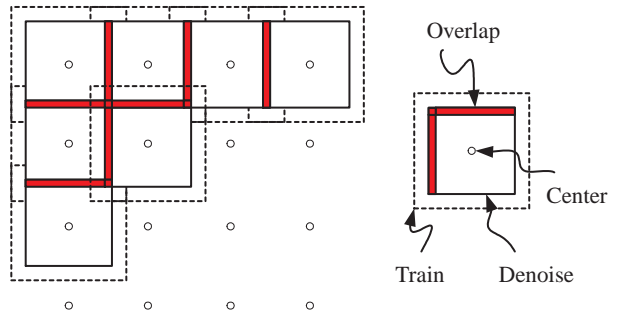


Fig. 1. Description of the denoising process. Denoising is done for each patch, labeled “Denoise” using the PC denoising algorithm. Training set for determining the principal components is generated from the patch labeled “Train”. The “Denoise” patches are overlapped in the dark region labeled “Overlap” in order to eliminate blocking artifacts.

decomposing the signal \mathbf{y}_i using the PC basis preserves the statistical independence and statistical properties of the coefficients n_i^l . Next, the PC coefficients x_i^l are modeled as realizations of $\mathcal{N}(0, \sigma_l^2)$ and for clarity the superscript l is dropped from the notation. The LMMSE estimator for x_i is:

$$\hat{x}_i = k \times y_i, \quad (4)$$

where k is constant. Forcing the orthogonality condition $E[(x_i - \hat{x}_i)y_i^T] = 0$ the estimate for k is:

$$k = \frac{E[x_i y_i]}{E[y_i^2]} = \frac{E[x_i(x_i + n_i)]}{E[(x_i + n_i)^2]}. \quad (5)$$

With n_i independent of x_i , $E[x_i n_i] = E[x_i]E[n_i] = 0$ and equation (5) becomes:

$$k = \frac{E[x_i^2]}{E[x_i^2] + E[n_i^2]} = \frac{\sigma_l^2}{\sigma_l^2 + \sigma^2} \quad (6)$$

and the LMMSE estimator of equation (4) is

$$\hat{x}_i = \frac{\sigma_l^2}{\sigma_l^2 + \sigma^2} \times y_i \quad (7)$$

The variance of x_i is estimated using the maximum likelihood estimator given in [4]:

$$\hat{\sigma}_l^2 = \max \left[0, \frac{1}{M} \sum_{i=1}^M y_i^2 - \sigma^2 \right], \quad (8)$$

where M is the number of training vectors in set S . (Recall that in equation (8), y_i is y_i^l but the superscript l is left out for clarity.)

Starting with a noisy image, the complete denoising algorithm is:

1. Estimate the noise variance, σ^2 , using equation (2).

2. Partition the image into overlapping patches as shown on the left side of Fig. 1. Each patch, depicted on the right side of Fig. 1, contains a *train region*, a *denoise region* and an *overlap region*. The *overlap region* is included in the *denoise region*, which is included in the *train region*.
3. Fix the dimension, N^2 , of the training vectors and generate S . The training vectors are $N \times N$ patches, reordered in an N^2 long vector, and the training set S is the collection of all the possible $N \times N$ patches included in the *train region*. To be consistent, the number of training vectors in S is M . The dimension of S , the matrix formed by ordering the vectors in S as column vectors, is $N^2 \times M$.
4. The PC basis functions are the eigenvectors of $\mathbf{Q} = (SS^T)^{-1}$, which are also the principal components of S .
5. For $l = 1 \dots N^2$ and $i = 1 \dots M$ find the PC coefficients y_i^l by taking projections of the training vectors in S onto the PC basis functions.
6. For $l = 1 \dots N^2$ estimate the variance of the PC coefficients using equation (8).
7. Denoise the PC coefficients using equation (7) and reconstruct the denoised training vectors in S . Since the training vectors in S overlap, average out the results in regions of overlap after the denoised training vectors are put back into the *train region*. In the middle of the *train region* each pixel is estimated N^2 times, while on the boundary, it may be estimated only once. Choose the *denoise region* such that each pixel is estimated N^2 times. This step resembles the over-complete basis denoising algorithm of [3]. The training vectors are formed from a moving window, which is similar to shifting the signal. In this sense the denoising algorithm has a built-in shift invariant feature.
8. If the *denoise region* is too large, blocking artifacts in the denoised image can become a problem, even though the PSNR values are still good. To average out the blocking artifacts between different denoised regions, add an *overlap region*.

3. RESULTS AND CONCLUSION

Using four different images, adaptive PC denoising is tested against three different algorithms. Three of the images used in our tests are the standard 512×512 gray level images of *lena*, *boat* and *barb*. The *ring* image is a 1-D chirp signal rotated around 360 degrees. It contains concentric circles

Image	Noisy	HMM	Sp. A.	SI-AShr	PC
Lena	14.13	26.07	27.23	27.96	28.00
Boat	14.15	24.70	25.74	26.29	26.36
Barb	14.16	23.29	25.03	24.72	26.21
Ring	14.15	17.58	23.28	22.97	27.18

Table 1. Comparison of denoising results. Values are peak-to-peak signal-to-noise ratio (PSNR) ($20 \log_{10}(255/\sigma_{error})$), with Gaussian noise of variance $\sigma = 50$.

Image	Noisy	HMM	Sp. A.	SI-AShr	PC
Lena	20.16	29.22	30.64	31.05	31.26
Boat	20.16	28.01	28.99	29.42	29.62
Barb	20.20	26.75	28.67	28.58	29.91
Ring	20.17	22.10	27.66	27.54	32.82

Table 2. Comparison of denoising results (PSNR, in dB) with Gaussian noise of variance $\sigma = 25$.

that get closer and closer to each other as they move outward, away from the center. All of these test images can be obtained from [10]. Each of the five images is corrupted with i.i.d. Gaussian noise of different variance, σ^2 , using the MATLAB command $+\sigma \times RANDN$.

The four algorithms used for testing are: the wavelet-domain Hidden Markov Tree Models of [8], the spatially adaptive image denoising algorithm of [5], the SI-Adaptive Shrink of [4], and our adaptive PC denoising. The adaptive PC denoising used a training region of 21×21 , a training vector size of 5×5 , a denoise region of 7×7 , and an overlap patch of 3. All tested algorithms were obtained directly from the authors and our tests matched closely the results the authors originally published. The results for σ values 50, 25 and 15 are in Tables 1, 2, and 3 respectively. In almost all experiments the adaptive PC denoising algorithm performed best in terms of PSNR values. Looking at the denoising results for $\sigma = 50$, the differences between the four denoising algorithms are very noticeable. The zoomed knee of *barb* is depicted in Fig. 2. Due to limited space, only the results of SI-AdaptShrink and adaptive PC are shown. From the four algorithms, PC denoising performs best at maintaining the high frequency contents of the stripes. On the test pattern *ring*, adaptive PC outperforms SI-AdaptShrink by about 5.0 dB. Images *ring* and *barb* are two images that emphasize the strength of our adaptive PC model. In particular, adaptive PC denoising performs best when there are strong edge patterns. The reader is also encouraged to visit [10, 11] for more image examples.

In conclusion, this paper presented a novel and simple approach to decomposing an image using adaptive principal components. The paper emphasized the strengths of this new decomposition approach by applying it to image de-

Image	Noisy	HMM	Sp. A.	SI-AShr	PC
Lena	24.60	31.66	33.05	33.31	33.60
Boat	24.59	30.68	31.65	32.01	32.25
Barb	24.63	29.39	31.49	31.56	32.63
Ring	24.61	25.80	30.97	31.07	36.32

Table 3. Comparison of denoising results (PSNR, in dB) with Gaussian noise of variance $\sigma = 15$.

noising.

4. ACKNOWLEDGMENTS

Many thanks go to Grace Chang, Xin Li, Hyeokho Choi, and Justin Romberg² for providing code of their white Gaussian denoising algorithms.

5. REFERENCES

- [1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [2] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. on Information Theory*, vol. 45, no. 7, pp. 2225–2237, 1999.
- [3] R. R. Coifman and D. L. Donoho, *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, New-York: Springer-Verlag, 1995.
- [4] S. Grace Chang, B. Yu, and Martin Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1522–1531, 2000.
- [5] Xin Li and Michael T. Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *IEEE Proc. ICIP*, 2000.
- [6] E. P. Simoncelli and W. T. Freeman., "The steerable pyramid: A flexible architecture for multi-scale derivative computation.," in *Second Int'l Conf on Image Proc.*, 1995, vol. 3, pp. 444–447.
- [7] Hoyer P.; Hyvarinen A.; Oja E., "Sparse code shrinkage for image denoising," *Neural Networks Proceedings. IEEE World Congress on Computational Intelligence.*, vol. 2, pp. 859–864, 1998.

²Hyeokho Choi and Justin Romberg's HMM denoising code is available on their web site.

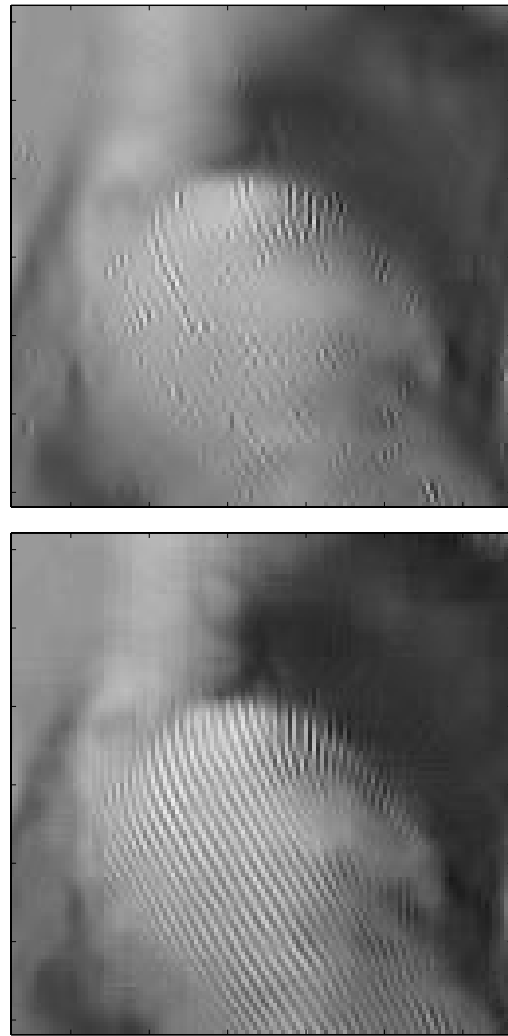


Fig. 2. Barbara's knee ($\sigma = 50$): SI-AdaptShrink (top) and adaptive PC (bottom). Notice how much better the stripes are preserved using adaptive PC.

- [8] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. On Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.
- [9] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of Royal Statist. Soc.*, vol. B 59, pp. 319–351, 1997.
- [10] DSP Lab at Cornell University, "<http://dsplab.ece.cornell.edu>," 2002.
- [11] Digital Multi Media Design (DMMD), "<http://www.dmmd.net/Research>," 2003.