

# PRINCIPAL COMPONENTS FOR NON-LOCAL MEANS IMAGE DENOISING

*Tolga Tasdizen*

Electrical and Computer Engineering Department, University of Utah

## ABSTRACT

This paper presents an image denoising algorithm that uses principal component analysis (PCA) in conjunction with the non-local means image denoising. Image neighborhood vectors used in the non-local means algorithm are first projected onto a lower-dimensional subspace using PCA. Consequently, neighborhood similarity weights for denoising are computed using distances in this subspace rather than the full space. This modification to the non-local means algorithm results in improved accuracy and computational performance. We present an analysis of the proposed method's accuracy as a function of the dimensionality of the projection subspace and demonstrate that denoising accuracy peaks at a relatively low number of dimensions.

**Index Terms**— Non-local means, principal component analysis, image denoising, image neighborhoods.

## 1. INTRODUCTION

Data-driven descriptions of structure are becoming increasingly important in image processing applications such as denoising, regularization and segmentation. One strategy is to use collections of nearby pixels, i.e. image neighborhoods, as a feature vector for representing local structure. Image neighborhoods are rich enough to capture the local structures of real images, but do not impose an explicit model. This representation has been used as a basis for image denoising [1, 2, 3, 4] and for texture image segmentation [5]. For both denoising and segmentation, it has been demonstrated that the accuracy of this strategy is on the same level as state-of-the-art methods in general and exceeds them in particular types of images such as those that have significant texture patterns. The drawback is the relatively high computational cost. The image neighborhood feature vector is typically high-dimensional, e.g. it is 49 dimensional if  $7 \times 7$  neighborhoods are used. Hence, the computation of similarities between feature vectors incurs a large computational cost. In this paper, we propose to project the image neighborhood vectors to a lower-dimensional space by principal component analysis (PCA). Then, the neighborhood similarity weights required for denoising are computed

from distances in this lower-dimensional space resulting in significant computational savings. Furthermore, in Section 4, we show that our approach results in increased accuracy over using the full image neighborhood vector.

The motivation for our approach stems from the assumption that image neighborhood vectors exist on a lower-dimensional manifold rather than the full space. This assumption is based on the observations by Huang and Mumford [6] and Lee *et al.* [7] who found multi-dimensional intensity data derived from image neighborhoods to be concentrated on low-dimensional manifolds. Even though these manifolds are unlikely to be linear, PCA can still be used to significantly reduce the dimensionality of image neighborhood vectors.

## 2. RELATED WORK

Buades *et al.* introduced the non-local means image denoising algorithm which averages the intensities of nearby pixels weighted by the similarity of image neighborhoods [1]. Image neighborhoods are typically defined as  $5 \times 5$ ,  $7 \times 7$  or  $9 \times 9$  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. More recently, Kervrann and Boulanger [3] have introduced an adaptive search-window approach which attempts to minimize the L2-risk with respect to the size of the search-window by analyzing the bias and variance of the estimator. Awate and Whitaker [2] introduced a statistical interpretation to the neighborhood-weighted averaging methods. Their approach is based on treating image neighborhoods as a random vector, computing its probability density function with non-parametric density estimation and formulating image denoising as an iterative reduction of that density.

Mahmoudi and Sapiro have proposed a method to improve the computational efficiency of the non-local means algorithm [8]. Their 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 [8] and other methods, i.e. Gabor filter responses, which use predetermined feature vector definitions

This work was supported by NSF CCF0732227 and NIH R01 EB005832. The author would also like to acknowledge the support of the Utah Science Technology and Research Initiative (USTAR) and the Scientific Computing and Imaging Institute.

to construct relatively lower-dimensional representations, the lower-dimensional vectors computed by PCA are data-driven and are approximations to the full neighborhood vector.

PCA of image neighborhoods was previously used for denoising [4]. However, in that work, PCA is computed for local collections of image neighborhood samples and denoising is achieved by direct modification of the principal components. In this paper, PCA is computed once, globally rather than locally which results in a more computationally efficient algorithm. Also, we use the non-local means averaging scheme rather than direct modification of principal components.

### 3. METHODS

Starting from a true, discrete image  $u$ , a noisy observation of  $u$  at pixel  $i$  is defined as  $v(i) = u(i) + n(i)$ . Let  $\mathcal{N}_k$  and  $\mathbf{v}(\mathcal{N}_k)$  denote a square neighborhood of fixed size centered around pixel  $k$  and the image neighborhood vector whose elements are the gray level values of  $v$  at  $\mathcal{N}_k$ , respectively. Also,  $\mathcal{S}_k$  is a square search-window of fixed size centered around pixel  $k$ . Then, the non-local means algorithm [1] defines an estimator for  $u$  at pixel  $i$  as

$$\hat{u}(i) = \sum_{j \in \mathcal{S}_i} \frac{1}{Z(i)} e^{-\frac{\|\mathbf{v}(\mathcal{N}_i) - \mathbf{v}(\mathcal{N}_j)\|^2}{h^2}} v(j), \quad (1)$$

where  $Z(i) = \sum_{j \in \mathcal{S}_i} e^{-\|\mathbf{v}(\mathcal{N}_i) - \mathbf{v}(\mathcal{N}_j)\|^2/h^2}$  is a normalizing term and parameter  $h$  controls the extent of averaging.

We propose to replace the distances  $\|\mathbf{v}(\mathcal{N}_i) - \mathbf{v}(\mathcal{N}_j)\|^2$  in (1) by distances computed from projections of  $\mathbf{v}(\mathcal{N})$  onto a lower-dimensional subspace determined by PCA. Let  $M$  be the number of pixels in the image neighborhood  $\mathcal{N}$ . Also, let  $\{\mathbf{b}_p\}_{p=1}^M$  be the eigenvectors of the  $M \times M$  empirical covariance matrix for the set of all image neighborhood vectors  $\{\mathbf{v}(\mathcal{N}_j)\}_{j=1}^Q$  where  $Q$  denotes the total number of pixels in the image. Furthermore, the eigenvectors are sorted in descending order according to their respective eigenvalues. Then, the projections of the image neighborhood vectors onto the  $d$ -dimensional PCA subspace is

$$\mathbf{v}[d](\mathcal{N}_i) = \sum_{p=1}^d \underbrace{\langle \mathbf{v}(\mathcal{N}_i), \mathbf{b}_p \rangle}_{f_p(\mathcal{N}_i)} \mathbf{b}_p, \quad (2)$$

where  $f_p(\mathcal{N}_i)$  is the length of  $i$ 'th vector's projection onto the  $p$ 'th basis vector. Due to the orthonormality of the basis

$$\|\mathbf{v}[d](\mathcal{N}_i) - \mathbf{v}[d](\mathcal{N}_j)\|^2 = \sum_{p=1}^d (f_p(\mathcal{N}_i) - f_p(\mathcal{N}_j))^2. \quad (3)$$

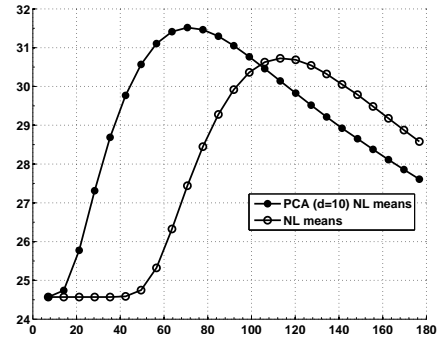
Finally, we define a new family of estimators for  $d \in [1, M]$

$$\hat{u}[d](i) = \sum_{j \in \mathcal{S}_i} \frac{1}{Z_d(i)} e^{-\frac{\sum_{p=1}^d (f_p(\mathcal{N}_i) - f_p(\mathcal{N}_j))^2}{h^2}} v(j), \quad (4)$$

where  $Z_d(i) = \sum_{j \in \mathcal{S}_i} e^{-\sum_{p=1}^d (f_p(\mathcal{N}_i) - f_p(\mathcal{N}_j))^2/h^2}$  is the new normalizing term. Note that  $\mathbf{v}[M](\mathcal{N}_i) = \mathbf{v}(\mathcal{N}_i)$ ; therefore, the proposed approach with  $d = M$  is equivalent to the standard non-local means, i.e.  $\hat{u}[M](i) = \hat{u}(i)$ .

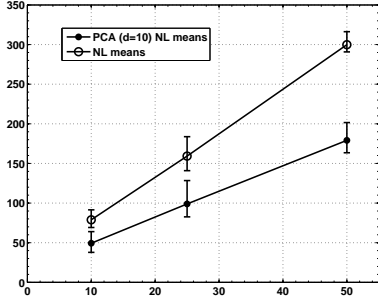
## 4. RESULTS

The proposed approach was tested on a set of eight images (shown in Figure 3) including those used in [9] and several additional images. Images were corrupted with additive, independent Gaussian noise and denoised using a  $7 \times 7$  image neighborhood  $\mathcal{N}$  and a  $21 \times 21$  search window  $\mathcal{S}$  as in [1]. Buades *et al.* [1] also suggest using  $h = 10\sigma$  where  $\sigma$  is the standard deviation of the additive noise. For the proposed method, the optimal choice for  $h$  depends on the dimensionality  $d$  of the PCA subspace. To illustrate this point, Figure 1 shows the PSNR after denoising as a function of  $h$  for an image that was corrupted with Gaussian noise ( $\sigma = 15$ ). Note that the peak PSNR is obtained at a lower  $h$  value for the proposed approach with  $d = 10$  than for the standard non-local means algorithm. This observation conforms to our expectations because distances computed in a subspace are necessarily smaller than distances computed in the full space.



**Fig. 1.** PSNR (dB) as a function of the parameter  $h$  for the *peppers* image. The PSNR for the noisy image is 24.6.

In general, the denoising parameter can be chosen with the rule  $h = \alpha(d)\sigma$ . We determine the scalar  $\alpha(d)$  in the following manner. For each test image, a noise level  $\sigma$  and a given PCA subspace dimensionality  $d$ , the optimal value for  $h$  was found empirically. Figure 2 shows the mean, minimum and maximum of the optimal  $h$  value over the set of test images as a function of  $\sigma$  for  $d = 10$  and for the non-local means algorithm ( $d = 49$ ). Notice that, for any given  $d$ , there is a linear relationship between the mean of optimal  $h$  and  $\sigma$ . Then,  $\alpha(d)$  is chosen as the slope of this linear relationship. Figure 3 shows the PSNR after denoising as a function of  $d$  for all test images using the  $h$  values chosen in this manner. Results are presented for three levels of noise standard deviation  $\sigma = 10, 25, 50$ . Recall that  $d = 49$  is identical to the non-local means algorithm (the right-most data point on each



**Fig. 2.** Optimal  $h$  value as a function of Gaussian noise standard deviation  $\sigma$ . The data points correspond to the mean of the optimal  $h$  value over the set of 8 test images while the bars demonstrate the minimum and maximum optimal  $h$ .

curve in Figure 3). In all cases, the proposed approach outperforms the standard non-local means algorithm. The best results are obtained at a relatively low PCA subspace dimensionality  $d$ . More specifically, for all images except *barbara*, choosing  $d = 6$  yields either the highest PSNR or a PSNR very close to the highest. For these images, the curves shown in Figure 3 have a characteristic shape: steeply increasing PSNR for  $d < 6$ , a knee around  $d = 6$  and gradually declining PSNR for  $d > 6$ . The increased accuracy at lower  $d$  values can be attributed to the observation that distances computed in the lower-dimensional space are likely to be more accurate than distance computed from the full-dimensional space because PCA discards the most irrelevant dimensions. The *barbara* image presents an exception where the best choice for  $d$  ranges from 10 to 15 depending on the amount of noise.

The computational complexity of the non-local means algorithm is  $O(QRM)$  where  $Q$ ,  $R$  and  $M$  are the number of pixels in the image, in the search window  $\mathcal{S}$  and in the neighborhood vector  $\mathcal{N}$ , respectively. In comparison, the complexity when using a  $d$ -dimensional subspace is  $O(QRd)$ . The additional costs in building the covariance matrix for PCA and computing the coefficients  $f_p$  in (2) are  $O(QM^2)$  and  $O(QMd)$ , respectively. Unlike [4], which denoises images directly by local PCA projections, our approach computes the PCA once globally. Eigenvectors are computed once globally for a small matrix ( $M \times M$ ); hence, this cost is negligible. Therefore, the total complexity for the proposed approach is  $O(Q(Rd + M^2 + Md))$ . This is significantly smaller than the non-local means algorithm cost because typically  $R \gg M$ . For the specific window sizes used in this work ( $R = 441$  and  $M = 49$ ), the non-local means cost is  $O(21609Q)$ . In comparison, the cost for our approach with  $d = 6$  is  $O(5341Q)$ . Furthermore, the covariance matrix can be estimated from a small fraction of the image neighborhood vectors, resulting in further computational savings. For instance, if 10% of the vectors are used for this purpose, the cost is reduced to  $O(3080Q)$ .

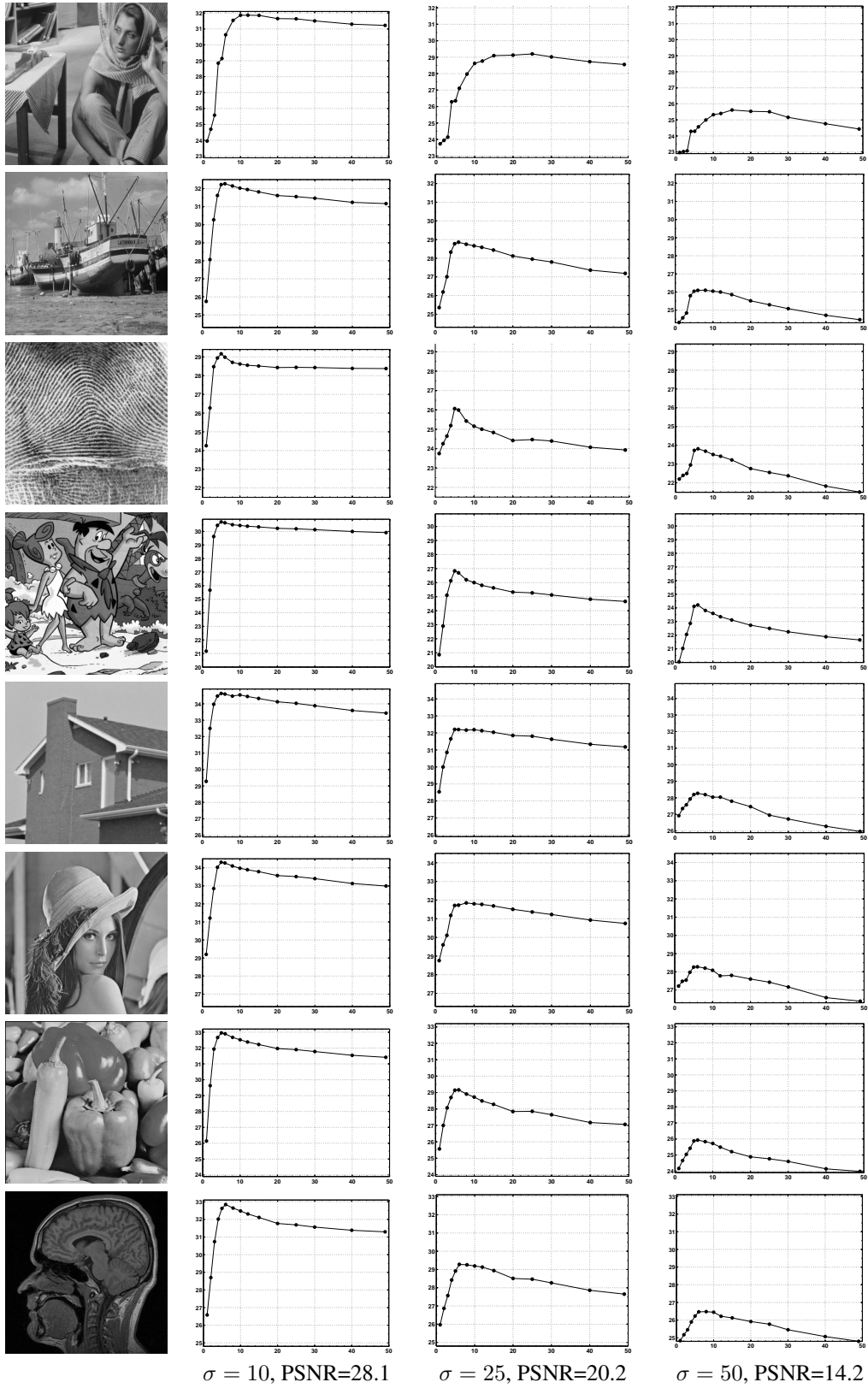
## 5. DISCUSSION

We showed that both the accuracy and computational cost of the non-local means image denoising algorithm [1] can be improved by computing neighborhood similarities after a PCA projection. Unlike the predetermined filters introduced in [8] for reducing the non-local means computational cost, our approach is data-driven and can adapt to the statistics of a given image. In [8] after the selection of neighborhoods to include in the weighted average, the weights are computed from the original high-dimensional vectors. In our approach, the lower-dimensional projections are not only used as a search criteria but also for computing neighborhood similarities resulting in increased accuracy in addition to reduced computational cost. Both approaches can also be easily applied to other denoising and segmentation algorithms that use similarity measures based on image neighborhood vectors [2, 3, 5].

We found that denoising accuracy peaked at  $d = 6$  for all except one test image suggesting that the choice of PCA subspace dimensionality can be fixed for a wide class of images. An interesting question is whether there is a fundamental difference in the complexity of image neighborhoods of the *barbara* image. As mentioned in Section 1, the manifold of image neighborhood vectors is unlikely to be linear. Hence, the number of dimensions for the subspace can potentially be further reduced by employing nonlinear dimensionality reduction methods instead of PCA.

## 6. REFERENCES

- [1] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *IEEE CVPR*, 2005, pp. 60–65.
- [2] S. P. Awate and R. T. Whitaker, "Higher-Order Image Statistics for Unsupervised, Information-Theoretic, Adaptive, Image Filtering," in *IEEE CVPR*, 2005, vol. 2, pp. 44–51.
- [3] C. Kervrann and J. Boulanger, "Unsupervised patch-based image regularization and representation," in *ECCV*, 2006, pp. 555–567.
- [4] D. D. Muresan and T. W. Parks, "Adaptive principal components and image denoising," in *IEEE ICIP*, 2003, vol. 1, pp. 101–104.
- [5] S. P. Awate, T. Tasdizen, and R. T. Whitaker, "Unsupervised Texture Segmentation with Nonparametric Neighborhood Statistics," in *ECCV*, 2006, pp. 494–507.
- [6] J. Huang and D. Mumford, "Statistics of natural images and models," in *ICCV*, 1999, pp. 541–547.
- [7] A. Lee, K. Pedersen, and D. Mumford, "The nonlinear statistics of high-contrast patches in natural images," *IJCV*, vol. 54, pp. 83–103, 2003.
- [8] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *IEEE Signal Processing Letters*, vol. 12, no. 12, pp. 839–842, 2005.
- [9] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. on Image Processing*, vol. 12, pp. 1338–1351, 2003.



**Fig. 3.** PSNR (dB) vs. PCA subspace dimensionality for three noise levels.