PATCH-BASED LOCALLY OPTIMAL DENOISING

Priyam Chatterjee and Peyman Milanfar

Department of Electrical Engineering
University of California, Santa Cruz
Email : {priyam, milanfar}@soe.ucsc.edu

ABSTRACT

In our previous work [1], we formulated the fundamental limits of image denoising. In this paper, we propose a practical algorithm where the motivation is to realize a locally optimal denoising filter that achieves the lower bound. The proposed method is a patch-based Wiener filter that takes advantage of both geometrically and photometrically similar patches. The resultant approach has a nice statistical foundation while producing denoising results that are comparable to or exceeding the current state-of-the-art, both visually and quantitatively.

Index Terms— Image denoising, image clustering, Wiener filter, LMMSE estimator, denoising bounds.

1. INTRODUCTION

The field of image denoising has seen a flurry of activity in recent years (see [1] and references therein) that have advanced the state-of-the-art considerably. Approaches proposed range from locally adaptive kernel-based methods [2], to non-local methods first introduced in [3]. Other methods perform denoising by modeling natural images, implicitly [4] or explicitly [5, 6]. One of the best performing methods, BM3D [7], uses a hybrid approach of combining spatial and transform (say, DCT) domain information to achieve state-of-the-art results. Although quite diverse in their approaches to denoising, the performances of many recent methods are surprisingly comparable. This motivated us to seek the fundamental limits of image denoising in [1]. Our findings there are used here to develop a practical algorithm that achieves essentially state-of-the-art denoising, while providing a solid statistical framework that justifies its performance.

In [1], we studied the problem of denoising from an estimation theoretic point of view. The problem was posed as that of estimating a patch $\mathbf{z}_i$ from its noise-corrupted observation

$$\mathbf{y}_i = \mathbf{z}_i + \boldsymbol{\eta}_i, \qquad i = 1, \ldots, M, \tag{1}$$

where $\boldsymbol{\eta}_i$ is assumed to be noise patches that are pixel-wise independent and identically distributed, and M is the number of such patches in the image. A Bayesian Cramér-Rao bounds framework [8] was used to derive a lower bound on the mean squared error (MSE) as

$$E\left[\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2\right] \geqslant \mathrm{Tr}\left[\left(N_i \frac{\mathbf{I}}{\sigma^2} + \mathbf{C}_{\mathbf{z}}^{-1}\right)^{-1}\right], \tag{2}$$

where $\hat{\mathbf{z}}_i$ is any estimate of $\mathbf{z}_i$. The expression above takes this specific form when the noise is assumed to be additive white Gaussian (AWGN) with covariance $\sigma^2\mathbf{I}$, but our analysis in [1] is more general.

Note that the formulation of (2) takes into account the geometric complexity of the patch as captured by the patch covariance $\mathbf{C}_{\mathbf{z}}$, where such complexity is independent of the actual patch intensity. The formulation also takes into account the level of redundancy N_i observed for each patch $\mathbf{z}_i$. Using such repeating patterns to improve performance forms the core of most popular denoising algorithms [3, 7]. This framework is in keeping with the intuition that the expected MSE increases with increasing patch complexity and noise variance. Bounds computed on various images in [1] indicate that modern denoising methods achieve near-optimal performance for images with high "semi-stochastic" content (that is, complicated image), whereas those with higher levels of redundancy, as observed typically in smoother images, could still be better denoised.

Apart from providing a formulation for bounding denoising performance, our analysis in [1] also provides a direction for realizing the optimal denoiser (in terms of MSE). Interest-

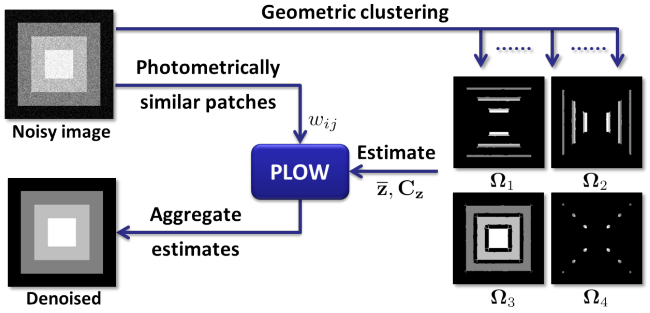


Fig. 1. Outline of our proposed patch-based locally optimal Wiener (PLOW) filtering method.

This work was supported in part by the AFOSR Grant FA9550-07-1-0365 and NSF Grant CCF-1016018.

ingly, for the case of AWGN, the bounds formulation in (2) is exactly the estimation error that can be expected from a linear minimum mean square error (LMMSE) estimator [8]. For the data model of Eq. 1 the LMMSE estimator is:

$$\widehat{\mathbf{z}}_i = \bar{\mathbf{z}} + \left(\mathbf{C}_{\mathbf{z}}^{-1} + \frac{\mathbf{I}}{\sigma^2}\right)^{-1} \frac{(\mathbf{y}_i - \bar{\mathbf{z}})}{\sigma^2}, \tag{3}$$

where $\bar{\mathbf{z}}$ and $\mathbf{C}_{\mathbf{z}}$ are the mean and covariance of $\mathbf{z}$. This is essentially the Wiener filter, popularly used in a transform domain [7]. However, the LMMSE estimator above does not take advantage of multiple photometrically similar patches that exist in most natural images. In this paper, we extend this patch-based spatial domain Wiener filter to exploit such patch redundancies to improve performance. Our method, described next, makes use of both geometrically, and photometrically, similar patches to determine the filter parameters.

The outline of our algorithm is illustrated in Fig. 1. First, we need to identify geometrically and photometrically similar patches. We cluster the image into regions of similar geometric structure, as shown in Fig. 1. From patches within each cluster, we estimate the moments $\bar{\mathbf{z}}$ and $\mathbf{C}_{\mathbf{z}}$. Photometrically similar patches are identified for each patch as the reference, from which we compute weights that ensure higher contribution for the most similar patches. The estimated parameters are then used to denoise each patch. These estimated patches are then combined to produce the denoised image. Next, we motivate and describe each of these steps in greater detail.

2. PATCH-BASED LOCALLY OPTIMAL WIENER FILTER (PLOW)

As mentioned earlier, the estimator in Eq. 3 can be improved by accounting for the presence of photometrically similar patches. Such similar patches need to be identified directly from the noisy image. We define a patch $\mathbf{y}_j$ to be *photometrically* similar to a reference $\mathbf{y}_i$ if it satisfies the condition

$$\|\widetilde{\boldsymbol{\varepsilon}}_{ij}\|^2 \leqslant \gamma_n^2 \quad \text{where} \quad \widetilde{\boldsymbol{\varepsilon}}_{ij} = \mathbf{y}_j - \mathbf{y}_i. \tag{4}$$

The threshold γ_n^2 is chosen such that it accounts for the level of noise corrupting the image patches [9]. We can then define an alternate data model for the entire set of patches that are similar to a reference patch $\mathbf{y}_i$ (inclusive) as

$$\begin{aligned} \underline{\mathbf{y}}_i = \mathbf{A}_i\mathbf{y}_i + \underline{\widetilde{\boldsymbol{\varepsilon}}}_i \quad &= \mathbf{A}_i(\mathbf{z}_i + \boldsymbol{\eta}_i) + (\underline{\boldsymbol{\varepsilon}}_i + \underline{\boldsymbol{\eta}}_i - \mathbf{A}_i\boldsymbol{\eta}_i) \\ = \mathbf{A}_i\mathbf{z}_i + \underline{\boldsymbol{\varepsilon}}_i + \underline{\boldsymbol{\eta}}_i \quad &= \mathbf{A}_i\mathbf{z}_i + \underline{\boldsymbol{\zeta}}_i. \end{aligned} \tag{5}$$

Here $\underline{\mathbf{y}}_i$ is a vector formed by concatenating all the $\mathbf{y}_j$ patches that are photometrically similar to $\mathbf{y}_i$, and $\mathbf{A}_i$ is the matrix formed by vertically stacking N_i identity matrices, each of size $n \times n$. $\underline{\boldsymbol{\eta}}_i$ denotes the corresponding noise patches $\boldsymbol{\eta}_j$ stacked together, and $\underline{\widetilde{\boldsymbol{\varepsilon}}}_i$ and $\underline{\boldsymbol{\varepsilon}}_i$ are vectors consisting of concatenated difference vectors $\widetilde{\boldsymbol{\varepsilon}}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}$, respectively, with

$$\boldsymbol{\varepsilon}_{ij} = \mathbf{z}_j - \mathbf{z}_i = \widetilde{\boldsymbol{\varepsilon}}_{ij} + \boldsymbol{\eta}_i - \boldsymbol{\eta}_j. \tag{6}$$

Since only similar patches are considered, $\boldsymbol{\varepsilon}_{ij}$ is small, resulting in $\underline{\boldsymbol{\zeta}}_i = \underline{\boldsymbol{\varepsilon}}_i + \underline{\boldsymbol{\eta}}_i$ exhibiting approximately Gaussian characteristics, even for mild noise.

The LMMSE estimator for the revised data model of Eq. 5 takes the form [8]:

$$\widehat{\mathbf{z}}_i = \bar{\mathbf{z}} + \left(\mathbf{C}_{\mathbf{z}}^{-1} + \mathbf{A}_i^T\mathbf{C}_{\underline{\boldsymbol{\zeta}}_i}^{-1}\mathbf{A}_i\right)^{-1} \mathbf{A}_i^T\mathbf{C}_{\underline{\boldsymbol{\zeta}}_i}^{-1}\left(\underline{\mathbf{y}}_i - \mathbf{A}_i\bar{\mathbf{z}}\right), \tag{7}$$

where $\mathbf{C}_{\underline{\boldsymbol{\zeta}}_i}$ denotes the covariance matrix for the error vector $\underline{\boldsymbol{\zeta}}_i$. When $\boldsymbol{\varepsilon}_i$ vectors are assumed to be independent of each other, we obtain a diagonal form for $\mathbf{C}_{\underline{\boldsymbol{\zeta}}_i}$ as

$$\mathbf{C}_{\underline{\boldsymbol{\zeta}}_i} = \mathbf{C}_{\underline{\boldsymbol{\varepsilon}}_i} + \mathbf{C}_{\underline{\boldsymbol{\eta}}_i} = \begin{bmatrix} \ddots & & \mathbf{0} \\ & \delta_{ij}^2\mathbf{I} & \\ \mathbf{0} & & \ddots \end{bmatrix}, \tag{8}$$

$$\begin{aligned} \text{where } \delta_{ij}^2 &= \frac{1}{n}E\left[\|\mathbf{z}_i - \mathbf{z}_j\|^2\right] + \sigma^2 \qquad (9) \\ &= \frac{1}{n}(E\left[\|\mathbf{y}_i - \mathbf{y}_j\|^2\right] - 2\sigma^2 n) + \sigma^2 \\ &= \frac{1}{n}E\left[\|\mathbf{y}_i - \mathbf{y}_j\|^2\right] - \sigma^2. \qquad (10) \end{aligned}$$

Substituting Eq. 8 in Eq. 7, and denoting $w_{ij} = \delta_{ij}^{-2}$, we obtain a simplified expression for our estimator as (refer [10] for derivation)

$$\widehat{\mathbf{z}}_i = \bar{\mathbf{z}} + \left(\mathbf{C}_{\mathbf{z}}^{-1} + \sum_{j=1}^{N_i} w_{ij}\mathbf{I}\right)^{-1} \sum_{j=1}^{N_i} w_{ij}\left(\mathbf{y}_j - \bar{\mathbf{z}}\right), \tag{11}$$

where it can be seen that N_i photometrically similar patches contribute to denoising $\mathbf{y}_i$, and the contributing weight of each $\mathbf{y}_j$ depends on the expected distance between the $\mathbf{y}_j$, $\mathbf{y}_i$ patch pair (Eq. 10). However, computing the *expected* distance from a single pair of $\mathbf{y}_i, \mathbf{y}_j$ observations is not practical. Consequently, we drop the expectation and enforce w_{ij} to be strictly non-negative through an alternate weight function

$$w_{ij} = \frac{1}{\sigma^2}\exp\left\{-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{h^2}\right\}, \tag{12}$$

where $h^2 = 1.75\sigma^2 n$ is an empirically chosen (but fixed) parameter that controls the amount of smoothing. The factor σ^{-2} ensures that when no photometrically similar patches are observed for any particular $\mathbf{y}_i$, $w_{ii} = \sigma^{-2}$ and the formulation of Eq. 11 coincides with Eq. 3. Also note that, without the expectation, Eq. 10 can be shown to be a first order Taylor approximation of Eq. 12 (see [10]).

The other parameters that need to be estimated from the noisy image are the moments $\bar{\mathbf{z}}$ and $\mathbf{C}_{\mathbf{z}}$. Considering intensity independent noise, denoising performance can be expected to be influenced by the complexity of patches captured by $\mathbf{C}_{\mathbf{z}}$, and not by the patch intensity. In keeping with this intuition, we follow the framework employed in [1, 9] and cluster the

image based on *geometric* similarity. The moments are, then, estimated from all patches within each cluster. An example of such geometric clustering is illustrated in Fig. 1 where a synthetic box image is divided into 4 representative clusters, each of which consists of patches containing either horizontal or vertical edges, smooth regions or corners.

To perform such clustering we use the locally adaptive regression kernels (LARK) [2] as features that capture the underlying patch structure. Initially developed for denoising [2], these features have been shown to be robust to noise, intensity and contrast changes among patches, making them ideal for geometric clustering [1, 9]. LARK features are computed for all the densely overlapping patches, and a K-Means clustering is performed to segment the image into K discontinuous geometric clusters. Within each cluster $\mathbf{\Omega}_k$, the patches are used to estimate the moments as

$$\hat{\bar{\mathbf{z}}} = E[\mathbf{y}_i \in \mathbf{\Omega}_k] \approx \frac{1}{M_k} \sum_{\mathbf{y}_i \in \mathbf{\Omega}_k} \mathbf{y}_i, \tag{13}$$

$$\hat{\mathbf{C}}_{\mathbf{z}} = \left[\hat{\mathbf{C}}_{\mathbf{y}} - \sigma^2 \mathbf{I} \right]_+, \tag{14}$$

where $\hat{\mathbf{C}}_{\mathbf{y}}$ is the sample covariance matrix, M_k is the number of patches in the k-th cluster, and $[\mathbf{X}]_+$ denotes the matrix $\mathbf{X}$ with its negative eigenvalues replaced by zero (or a very small positive value) [9]. Since the condition for geometric similarity is more relaxed compared to that for photometric similarity, we can expect more patches to lie within each cluster, resulting in a stable estimate of the moment parameters.

Once the parameters are estimated from the noisy image, they can be used in Eq. 11 to obtain a denoised estimate for each patch in the image. These patches being densely overlapped, we obtain multiple estimates for each pixel. To form the denoised image, we need to combine these multiple estimates to obtain a single final estimate for each pixel in the image. In the absence of any prior information about z_i, we use a weighted averaging scheme

$$\hat{z}_i = \sum_{r=1}^{R} \frac{v_{rl}^{-1} \hat{z}_{rl}}{\sum_r v_{rl}^{-1}}, \tag{15}$$

where the weight for each estimate of a pixel z_i, denoted by $\hat{z}_{rl}$ as the l-th pixel in the r-th patch, is the inverse of that pixel's estimate variance v_{rl}. This variance v_{rl} is the l-th diagonal entry of the error covariance of the r-th patch estimate:

$$\mathbf{C}_e \approx \left(\mathbf{C}_{\mathbf{z}}^{-1} + \sum_{j=1}^{N_i} w_{rj} \mathbf{I} \right)^{-1}. \tag{16}$$

Note that in this aggregation step only those (R) patches that contain pixel z_i are considered. This provides us with a final denoised image. We compare the performance of our proposed denoising algorithm with some recently proposed methods next.

3. RESULTS

In this section we compare our method to other popular denoising methods. In our approach, outlined in Fig. 1, we need to cluster the image into K segments. Using a large K results in too few patches being grouped together leading to possible inaccuracies in moment estimation. On the other hand, too few clusters may result in structurally dissimilar patches being grouped together. While K can be tuned for best performance on a given image, we found using $K = 15$ achieves close to the best denoising performance for our method.

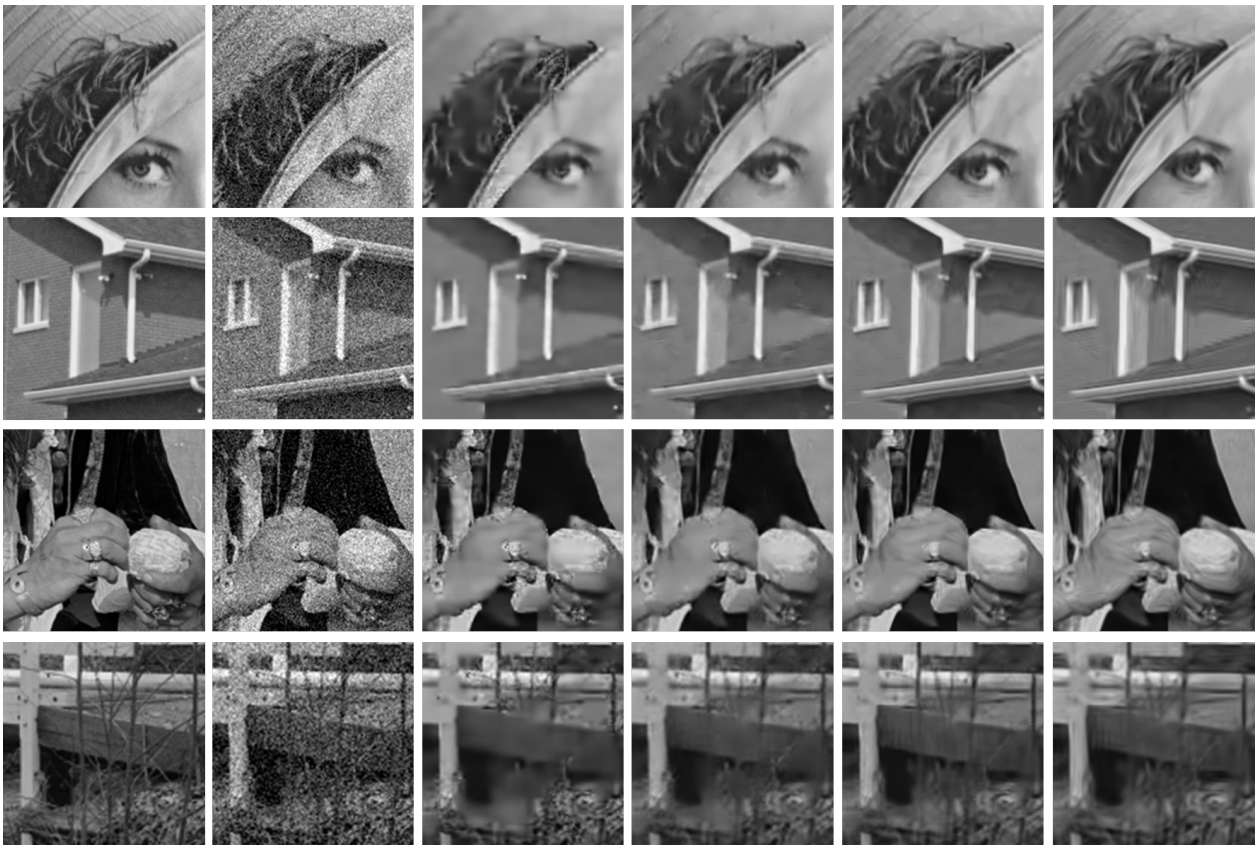
Another parameter that influences the performance is the patch size, which was empirically chosen to be 11×11. Our method also requires us to identify photometrically similar patches. In the interest of reducing execution time, we restrict ourselves to searching for such similar patches within a 30×30 pixel neighborhood of each reference patch. All our experiments were performed with these fixed parameters without the need for any manual tuning. This makes our approach practical.

The results obtained by our method are visually compared to some popular denoising methods in Fig. 2, where we show (cropped) results obtained for different images strongly corrupted by AWGN with $\sigma = 25$. Note how our method is capable of removing the noise while retaining much of the finer details, even for such high noise cases. Visually, our method can be seen to be superior to FoE [6] and K-SVD [4], while comparable to the state-of-the-art BM3D [7]. More such comparisons with different images and noise levels can be viewed at `http://users.soe.ucsc.edu/~priyam/PLOW/`.

We also provide quantitative evaluation of our method using peak signal-to-noise ratio (PSNR) as the performance metric. The PSNRs obtained by the methods over different noise levels are shown in Table 1. It can be seen there that our method consistently produces results that are also quantitatively comparable to the state-of-the-art.

Table 1. Quantitative comparison of denoising performance. PSNR values reported are the mean results from 5 different noise realizations for each case. PSNR is measured in dB as $10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right)$.

Image	σ	**FoE** [6]	**K-SVD** [4]	**BM3D** [7]	**PLOW**
Lena 512×512	5	38.24	38.55	38.73	38.66
	15	33.28	33.71	34.26	33.90
	25	30.83	31.28	32.07	31.92
House 256×256	5	38.23	39.33	39.80	39.52
	15	33.48	34.30	34.95	34.72
	25	31.11	32.12	32.89	32.70
Man 512×512	5	36.97	37.05	37.28	37.02
	15	30.22	30.46	30.98	30.26
	25	27.37	27.59	28.29	28.08
Stream 512×512	5	35.57	35.58	35.75	35.59
	15	28.41	28.51	28.74	28.71
	25	25.65	25.84	26.14	26.20



(a) Original (b) Noisy ($\sigma = 25$) (c) FoE [6] (d) K-SVD [4] (e) BM3D [7] (f) PLOW

Fig. 2. Comparison of denoising performance. Full images can be viewed at `http://users.soe.ucsc.edu/~priyam/PLOW/`.

4. CONCLUSIONS

In this paper, we proposed PLOW – a *non-local* patch-based LMMSE filter that takes advantage of patch redundancies. The method is designed with the motivation of achieving near optimal performance. The proposed algorithm has a sound theoretical framework that explains its denoising performance. Experimentally, we showed that our method performs comparably to the current state-of-the-art, both visually and in terms of PSNR.

The approach proposed here was applied to gray-scale images. However, the framework can be extended to denoising color images where inter-channel correlations are taken into account. For this we need to estimate the covariance matrix for the full color patches, whereby the dependencies across color channels can be captured implicitly. This is a possible direction for future research.

5. REFERENCES

[1] P. Chatterjee and P. Milanfar, "Is denoising dead?" *IEEE Trans. on Img. Proc.*, vol. 19, no. 4, pp. 895–911, Apr. 2010.

[2] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. on Img. Proc.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.

[3] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising methods, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[4] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Img. Proc.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[5] S. Lyu and E. P. Simoncelli, "Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 693–706, Apr. 2009.

[6] S. Roth and M. J. Black, "Fields of experts," *Intl. J. Comp. Vis. (IJCV)*, vol. 82, no. 2, pp. 205–229, Apr. 2009.

[7] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. on Img. Proc.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, ser. Signal Processing. Upper Saddle River, N.J.: Prentice-Hall, Inc., 1993, vol. 1.

[9] P. Chatterjee and P. Milanfar, "Practical bounds on image denoising: From estimation to information," *IEEE Trans. on Img. Proc.*, vol. 20, no. 5, pp. 1221–1233, May 2011.

[10] ——, "Patch-based near-optimal image denoising," *IEEE Trans. on Img. Proc.*, 2011, submitted.