

# Image Denoising via Adaptive Soft-Thresholding Based on Non-Local Samples

Hangfan Liu, Ruiqin Xiong, Jian Zhang and Wen Gao \*  
Institute of Digital Media, Peking University, Beijing 100871, China

Email:{liuhf, rqxiong, jian.zhang, wgao}@pku.edu.cn

## Abstract

*This paper proposes a new image denoising approach using adaptive signal modeling and adaptive soft-thresholding. It improves the image quality by regularizing all the patches in image based on distribution modeling in transform domain. Instead of using a global model for all patches, it employs content adaptive models to address the non-stationarity of image signals. The distribution model of each patch is estimated individually and can vary for different transform bands and for different patch locations. In particular, we allow the distribution model for each individual patch to have non-zero expectation. To estimate the expectation and variance parameters for the transform bands of a particular patch, we exploit the non-local correlation of image and collect a set of similar patches as data samples to form the distribution. Irrelevant patches are excluded so that this non-local based modeling is more accurate than global modeling. Adaptive soft-thresholding is employed since we observed that the distribution of non-local samples can be approximated by Laplacian distribution. Experimental results show that the proposed scheme outperforms the state-of-the-art denoising methods such as BM3D and CSR in both the PSNR and the perceptual quality.*

## 1. Introduction

Image denoising is an important problem in many image processing tasks. It has attracted a lot of research interest in the past few decades [1–22]. It aims to recover the original image signal from its observed noisy version, which can be formulated as

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

where  $\mathbf{x}$  is the desired original image,  $\mathbf{n}$  represents the additive noise, and  $\mathbf{y}$  is the corrupted observation. To solve such

\*Corresponding author: R. Xiong. This work was supported in part by the National Natural Science Foundation of China (61370114, 61421062), Beijing Natural Science Foundation (4132039), Research Fund for the Doctoral Program of Higher Education (20120001110090) and also by the Cooperative Medianet Innovation Center.

ill-posed problem, it is critical to exploit the prior knowledge that characterizes the statistical features of the images.

Early regularization techniques mainly utilize the local correlation among image pixels. A typical example of this kind is the total variation (TV) regularization [9, 11]. It depicts the feature that natural images are smooth in most regions, and can be seen as a kind of sparsity regularization in gradient domain. Typical sparsity regularization models are based on the assumption that images can be sparsely presented in transform domain, *e.g.* discrete cosine transform (DCT) domain or discrete wavelet transform (DWT) domain [4, 5, 10, 14, 15]. These transforms are orthogonal or nearly orthogonal, and use fixed transform basis. However, it is impossible to efficiently present all kinds of patterns in images using a fixed basis. Therefore, [6] advocates to use content adaptive basis with the assistance of principal component analysis (PCA). Some later works use over-complete dictionary to represent the image signal and try to learn an optimal dictionary so that the sparsity of the signal is maximized [3, 23].

Since the invention of non-local means (NLM) denoising [7] which exploits the repetitiveness of patch patterns in image signals, extensive research works are motivated to take advantage of non-local similarity for image restoration tasks, and achieve superior performance over local regularization [2, 8, 12, 18–20, 22, 24–26]. Among these non-local similarity based schemes, the famous benchmark BM3D [2] is basically a combination of DCT coefficient thresholding and nonlocal block matching. It stacks similar blocks of a reference patch into a three dimensional (3D) block, on which 3D transform is applied, and hard thresholding (in the first step) or Wiener filtering (in the second step) is performed. Dong *et al.* [13] also proposed a two-stage denoising scheme but adopts PCA instead of DCT for local decorrelation. The so-called LPG-PCA method uses non-local similar patches as data samples to estimate statistical parameters for PCA training. Combining the ideas of structural clustering and dictionary learning, CSR [8] attempts to unify local and nonlocal sparsity constraints.

This paper aims to develop an effective image denoising scheme using adaptive signal modeling and adaptive

soft-thresholding. It regularizes all image patches based on adaptive distribution modeling in the transform domain. Instead of using a global model, the scheme employs content adaptive models that are estimated for each patch individually. In particular, we considered the possibility of non-zero expectation in the distribution of transform coefficient. To estimate the expectation and variance for each coefficient, we exploit non-local correlation and collect a set of similar patches as data samples to form the distribution. We use PCA to determine the decorrelation transform, and treat different transform bands separately according to their statistical characteristics. To study the distribution of non-local sample patches, we considered generalized Gaussian distributions (GGD) and observed that the distribution of non-local similar patches can be well approximated by Laplacian distribution. Based on the observation, adaptive soft-thresholding is adopted to implement the regularization.

The remainder of the paper is organized as follows. Section 2 briefly reviews the methods of image denoising from the local and nonlocal perspectives, respectively. Section 3 discusses how to model the distribution of transform coefficients adaptively. Section 4 describes the proposed image denoising scheme in details and Section 5 shows how the optimization problem can be efficiently solved. Experimental results are reported in Section 6 and Section 7 concludes the paper.

## 2. Background

The regularization techniques for image denoising problems can generally be divided into two categories: local regularization and non-local regularization. This section briefly reviews the basic ideas from these two perspectives. Just as most recent methods, this paper considers patch based denoising, which divides the image into overlapping patches and performs denoising on each patch, and then reconstructs the overall image by averaging the denoised patches. We denote the vectorized patch of size  $S \times S$  at location  $i$  by  $\mathbf{x}_i$ .

Local regularization utilizes the feature of image signals that they can be sparsely represented in transform domain:

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \sum_i \|\Phi \mathbf{x}_i\|_p^q, \quad (2)$$

where  $p$  is often set to be 0, 1 or 2,  $\mu$  is the regularization parameter controlling the trade-off between the two terms.  $\Phi$  is the transform matrix, which is able to decorrelate and sparsely represent the image signal. Since noise is usually independently distributed, it can be separated from the signal components after decorrelation and then suppressed via hard-thresholding (when  $p = 0$ ), soft-thresholding ( $p = 1$ ) or Wiener filtering ( $p = 2$ ).

Due to the existence of self-repeating patterns in image signals, nonlocal similar patches can act as data samples to

provide statistical information for the estimation of original signals. Furthermore, making use of nonlocal correlation often leads to sparser representation in transform domain, hence can distinguish signal components from noise more effectively and result in better denoising performance. The impressive success of nonlocal means (NLM) denoising [7] triggered a flurry of research works to utilize the nonlocal similarity of natural images. Different nonlocal denoising schemes utilize nonlocal correlation in various ways. For instance, NLM generates the estimated pixels by linear weighted average of nonlocal pixels in spatial domain, BM3D [2] enhances the sparsity of coefficients by an additional transform along the third dimension of three dimensional (3D) patch groups, LPG-PCA [13] treats nonlocal similar patches as samples, while CSR [8] gathers similar patches into a cluster and approximate them to the centroid.

## 3. Regularization via Adaptive Modeling

This paper introduces an adaptive denoising method based on transform domain adaptive distribution modeling. The key points of the proposed scheme include: 1) The regularization is applied based on distribution modeling in transform domain. 2) Instead of using a global model that assumes the same distribution for all transform coefficients, we employ adaptive models that are estimated for each patch individually and can vary for different locations and different transform bands. 3) Both the expectation and variance of the distribution are adaptively estimated. In particular, we consider the possibility that each individual coefficient may have **non-zero** expectation. 4) To estimate the distribution parameters for a particular patch, we use a set of non-local similar patches as the data samples to form the distribution. Dissimilar patches are excluded so that the parameters are not affected by irrelevant image contents. 5) The adaptive regularization is realized by adaptive soft-thresholding and overlapped patches are aggregated and averaged to generate the ultimate estimated image.

### 3.1. Adaptive Regularization

We may interpret Eq. (2) as the maximum a posteriori (MAP) estimation deduced in Bayesian framework, then the  $l_p$ -norm indicates the assumed distribution of the coefficients. Just as BM3D and LPG-PCA, we adopt  $l_2$ -norm in this section, which implicitly assumes the coefficients to be Gaussian distributed.

It is straightforward to deal with different bands equally. In this case, we denote the variance of the whole coefficients by scalar  $\sigma^2$  and the variance of noise by  $\sigma_n^2$ , then the MAP estimate of the original signal writes:

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2\sigma_n^2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{1}{2\sigma^2} \sum_i \|\Phi \mathbf{x}_i\|_2^2. \quad (3)$$

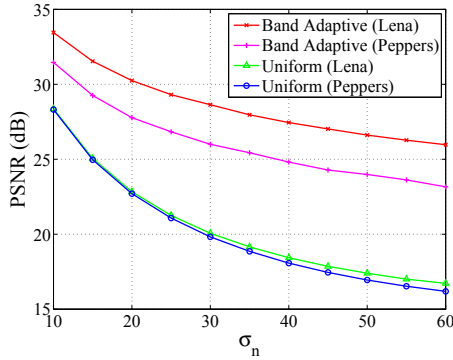


Figure 1. PSNR comparison for uniform denoising and band adaptive denoising. The input images are contaminated by Gaussian white noise, with standard deviation  $\sigma_n$  ranging from 10 to 60.

Such minimization problem leads to a *uniform* operation on all the coefficients.

The above consideration ignores the fact that statistical characteristics of coefficients in different bands may vary dramatically. Evidently, variance of coefficients in a low frequency band is usually much more significant than that in a high frequency band. For this reason, it is more plausible to treat different bands separately. This would make for *band adaptive* operation, which assigns coefficients with appropriate parameters adaptively. Let  $\vec{\sigma}_i$  be the standard deviation vector of  $\mathbf{x}_i$ , with the  $k$ -th element in  $\vec{\sigma}_i$  being the standard deviation of the  $k$ -th element in patch  $\mathbf{x}_i$ .  $\vec{\sigma}_i$  is of the same size as  $\mathbf{x}_i$ . In this setting, the MAP estimate of  $\mathbf{x}$  is formulated as:

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2\sigma_n^2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{1}{2} \sum_i \left\| \frac{\Phi \mathbf{x}_i}{\vec{\sigma}_i} \right\|_2^2, \quad (4)$$

where the division operation is component wise.

To verify the conjecture that the band adaptive scheme performs better, we compare the denoising results of Eq. (3) and Eq. (4), with  $\sigma_n$  ranging from 10 to 60. Fig. 1 shows the denoising results of images Lena and Peppers. Other images exhibit similar results. It is evident that the band adaptive parameter estimation would lead to much better denoising performance. Such observations encourage us to treat the transform coefficients adaptively according to their statistical characteristics. This paper uses PCA as the transform, which is known for its virtue of being signal adaptive and near optimal for decorrelation.

### 3.2. Distribution Modeling of Transform Coefficients

Just as most existing image restoration regularization, the formulation of Eq. (4) simply assumes the expectations of coefficients to be zero, which may not be accurate. In addition to variances, expectations of coefficients in different bands are also estimated adaptively in this paper.

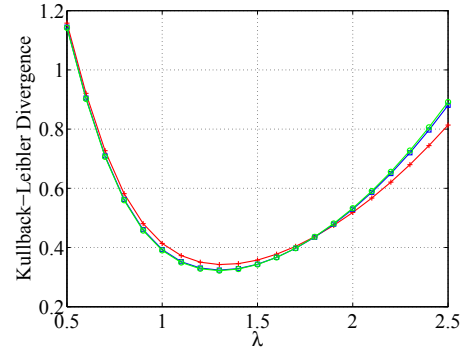


Figure 3. KL divergences between the distributions of coefficients and GGD with shape parameter  $\lambda$  ranging from 0.5 to 2.5. The three curves correspond to coefficients in three different bands. The optimal  $\lambda$  is near 1.3.

To model the distribution of coefficients, we fit the distribution of coefficients by general exponential family distributions, *i.e.* generalized Gaussian distribution (GGD), which is widely used in image coding and processing. The Generalized Gaussian density has the following form:

$$\mathcal{G}(u; \lambda) = \frac{1}{2\Gamma(1 + \frac{1}{\lambda})} \exp(-|u|^\lambda), \quad (5)$$

where  $\Gamma(\cdot)$  is the gamma function, and  $\lambda$  is the shape parameter which controls the overall shape of the distribution.  $\mathcal{G}(u; \lambda)$  is a Laplacian distribution when  $\lambda = 1$  and is a Gaussian distribution when  $\lambda = 2$ .

We extract more than  $6.5 \times 10^4$  groups from 9 natural images, with every group consisting of 60 similar patches, and apply PCA to each group. For the  $k$ -th band of the  $i$ -th group, suppose  $\mu_i^k$  and  $\sigma_i^k$  are respectively the mean value and standard deviation of the coefficients, then coefficients of this band (denoted by  $\mathbf{C}_i^k$ ) are centralized and variance-normalized:

$$\mathbf{C}_i^k := (\mathbf{C}_i^k - \mu_i^k) / \sigma_i^k. \quad (6)$$

After that, coefficients of the band  $k$  from all the groups are gathered together as samples and form the variance-normalized distribution, as shown in Fig. 2.

To find out the shape parameter  $\lambda$  that can best fit the distribution of PCA coefficients, we employ Kullback-Leibler (KL) divergence to measure the fitting error of GGD, with  $\lambda$  ranging from 0.5 to 2.5 for every band, as displayed in Fig. 3. The optimal  $\lambda$  falls near 1.3 for all the bands. For convenience of calculation, this paper sets  $\lambda = 1$ , *i.e.* the optimal GGD is approximated by Laplacian distribution.

## 4. The Denoising Scheme via Adaptive Soft-Thresholding

The proposed denoising scheme is based on the discussions in Section 3. Specifically speaking, the input image is

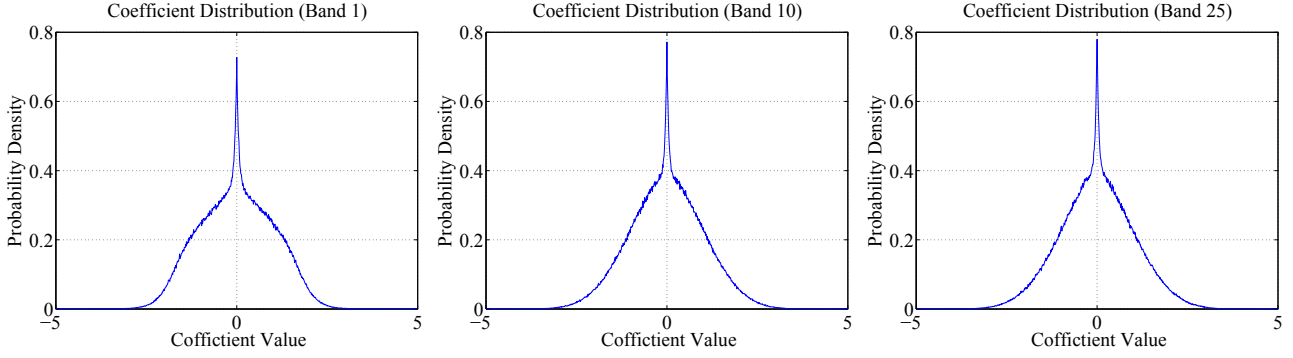


Figure 2. Variance-normalized distributions of 3 different bands.

divided into overlapping patches of size  $S \times S$ . For the current patch  $\mathbf{y}_i$ , we search non-locally for its similar patches. The dissimilarity between two patches  $\mathbf{y}_i$  and  $\mathbf{y}_j$  is measured by Euclidean distance:

$$d(i, j) = \frac{\|\mathbf{y}_i - \mathbf{y}_j\|_2^2}{S^2}. \quad (7)$$

Then the patches with distances smaller than a certain threshold  $\tau$  are considered to be ‘‘similar’’ and stacked into group  $Y_i$ :

$$Y_i = \{\mathbf{y}_j | d(i, j) \leq \tau\}, \quad (8)$$

and the corresponding locations are recorded in set  $\mathbb{L}_i$ :

$$\mathbb{L}_i = \{j | d(i, j) \leq \tau\}. \quad (9)$$

As an alternative, we may also group the  $M$  most similar patches into  $Y_i$ .

Besides, the similar patches of  $\mathbf{y}_i$  are used to train the PCA transform  $\Phi_i$ . Suppose  $\mathbf{p}_i^1, \mathbf{p}_i^2, \dots, \mathbf{p}_i^N$  are the selected training samples and let matrix  $P_i = [\mathbf{p}_i^1, \mathbf{p}_i^2, \dots, \mathbf{p}_i^N]$ . We calculate  $\bar{\mathbf{p}}_i$  by averaging every row of  $P_i$  so that the centralized version of  $P_i$  is estimated as  $\tilde{P}_i = [\mathbf{p}_i^1 - \bar{\mathbf{p}}_i, \mathbf{p}_i^2 - \bar{\mathbf{p}}_i, \dots, \mathbf{p}_i^N - \bar{\mathbf{p}}_i]$ . The covariance matrix of  $P_i$  is written as  $\text{Cov}(P_i) = \Psi_i \Lambda_i \Psi_i^T$ , where  $\Psi_i$  is the orthonormal eigenvector matrix and  $\Lambda_i$  is the sorted diagonal eigenvalue matrix. Then we adopt  $\Psi_i^T$  as the transform matrix  $\Phi_i$ .

Apply the learned transform  $\Phi_i$  to all the patches in  $Y_i$ . Let  $\vec{\mu}_i$  and  $\vec{\sigma}_i$  be the expectation vector and standard deviation vector of the  $\mathbf{x}_i$ , where the  $k$ -th entry of  $\vec{\mu}_i$  and  $\vec{\sigma}_i$  are the expectation and standard deviation of coefficients in band  $k$  respectively. Then the proposed objective function of this paper writes:

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{c}{\sigma_n^2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \sum_i \left\| \frac{\Phi_i \mathbf{x}_i - \vec{\mu}_i}{\vec{\sigma}_i} \right\|_1, \quad (10)$$

where  $c = \frac{1}{2\sqrt{2}}$ .

In practice, the expectation and standard deviation of clean data is not available. The expectation  $\vec{\mu}_i(k)$  is estimated as the median value of the coefficients in the  $k$ -th

band because the coefficients are assumed to conform to Laplacian distribution. The advantage of such estimation is to exclude influences of outliers. Furthermore, since the noise is assumed to be i.i.d. Gaussian distributed, the standard deviation of coefficients in the  $k$ -th band  $\vec{\sigma}_i(k)$  can be estimated by

$$\vec{\sigma}_i(k)^2 = \max(\vec{\sigma}_{y_i}(k)^2 - \sigma_n^2, 0), \quad (11)$$

where  $\vec{\sigma}_{y_i}$  is the standard deviation vector calculated by coefficients of observed noisy image  $\mathbf{y}$ . Let  $\vec{\beta}_j = \Phi_i \mathbf{y}_j$ ,  $j \in \mathbb{L}_i$ , then  $\vec{\sigma}_{y_i}$  is calculated as

$$\vec{\sigma}_{y_i}^2 = \frac{1}{M} \sum_{j \in \mathbb{L}_i} \left( \vec{\beta}_j - \vec{\mu}_i \right)^2, \quad (12)$$

where  $M$  is the number of similar patches in the group and the calculations are element wise.

Since the solution to this optimization problem can be boiled down to coefficient wise soft-thresholding operations, and the parameters are adaptively estimated by non-locally searched samples, we name this scheme as adaptive soft-thresholding based on non-local samples (AST-NLS).

## 5. Numerical Solutions

The frequencies of pixels appearing in the overlapping patches are roughly equal, hence

$$\|\mathbf{y} - \mathbf{x}\|_2^2 \approx c_1 \sum_i \|\mathbf{y}_i - \mathbf{x}_i\|_2^2, \quad (13)$$

here  $c_1$  is a positive constant. Let  $\vec{\alpha}_i = \Phi_i \mathbf{x}_i$ ,  $\vec{\beta}_i = \Phi_i \mathbf{y}_i$ . Considering the unitary property of  $\Phi_i$ , we have

$$\|\mathbf{y}_i - \mathbf{x}_i\|_2^2 = \|\Phi_i \mathbf{y}_i - \Phi_i \mathbf{x}_i\|_2^2 = \|\vec{\beta}_i - \vec{\alpha}_i\|_2^2, \quad (14)$$

thus we may write Eq. (10) as:

$$\vec{\alpha} = \arg \min_{\vec{\alpha}} \sum_i \frac{c_2}{\sigma_n^2} \|\vec{\beta}_i - \vec{\alpha}_i\|_2^2 + \sum_i \left\| \frac{\vec{\alpha}_i - \vec{\mu}_i}{\vec{\sigma}_i} \right\|_1, \quad (15)$$

Table 1. PSNR Comparison (Unit: dB)

$\sigma_n$	10				20				30			
Schemes	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>
Airplane	35.86	35.57	35.93	<b>36.14</b>	32.53	32.06	32.62	<b>32.81</b>	30.78	30.23	30.88	<b>31.09</b>
Barbara	34.53	34.99	35.02	<b>35.20</b>	30.89	30.94	31.20	<b>31.43</b>	28.64	28.46	28.83	<b>29.09</b>
C.man	33.92	33.61	33.93	<b>34.24</b>	30.15	29.73	30.29	<b>30.51</b>	28.27	27.83	28.54	<b>28.63</b>
Lena	35.05	35.01	35.31	<b>35.40</b>	31.29	31.07	31.58	<b>31.59</b>	29.20	28.79	29.44	<b>29.48</b>
Monarch	33.93	34.09	34.51	<b>34.92</b>	30.08	30.09	30.67	<b>30.89</b>	27.94	27.79	28.51	<b>28.70</b>
R.R.Hood	34.97	34.88	35.17	<b>35.28</b>	32.00	31.72	32.19	<b>32.23</b>	30.40	30.09	<b>30.63</b>	30.62
Sailboats	36.31	36.16	36.38	<b>36.47</b>	32.88	32.49	32.92	<b>33.06</b>	30.88	30.43	31.01	<b>31.11</b>
Window	36.51	36.31	36.85	<b>36.99</b>	32.67	32.16	32.89	<b>33.08</b>	30.49	29.76	30.63	<b>30.92</b>
Baboon	29.43	29.40	29.37	<b>29.53</b>	25.45	25.39	25.52	<b>25.65</b>	23.58	23.49	23.78	<b>23.88</b>
Couple	33.87	33.55	33.91	<b>34.02</b>	30.60	30.05	30.59	<b>30.68</b>	28.66	27.94	28.62	<b>28.77</b>
F.boat	33.75	33.62	33.83	<b>33.96</b>	30.62	30.24	30.72	<b>30.76</b>	28.77	28.26	28.89	<b>28.92</b>
House	36.39	36.17	<b>36.82</b>	36.77	33.39	33.05	33.87	<b>33.92</b>	31.56	31.12	32.07	<b>32.42</b>
Peppers	34.54	34.11	34.69	<b>34.88</b>	31.07	30.58	31.26	<b>31.36</b>	28.97	28.47	29.24	<b>29.35</b>
straw	30.78	31.35	31.49	<b>31.50</b>	26.92	27.08	<b>27.42</b>	27.35	24.76	24.55	25.03	<b>25.09</b>
<b>Average</b>	34.27	34.20	34.51	<b>34.66</b>	30.75	30.47	30.98	<b>31.09</b>	28.78	28.37	29.01	<b>29.15</b>
$\sigma_n$	50				70				90			
Schemes	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>
Airplane	28.67	27.94	28.60	<b>28.95</b>	27.27	26.46	27.17	<b>27.65</b>	26.19	25.39	26.13	<b>26.67</b>
Barbara	26.11	25.53	26.17	<b>26.27</b>	24.44	23.80	24.59	<b>24.63</b>	23.29	22.60	23.49	<b>23.63</b>
C.man	26.03	25.46	26.13	<b>26.31</b>	24.42	23.84	24.46	<b>24.74</b>	23.27	22.65	23.24	<b>23.59</b>
Lena	26.83	26.07	26.90	<b>27.13</b>	25.27	24.31	25.22	<b>25.58</b>	24.06	22.98	23.99	<b>24.44</b>
Monarch	25.57	24.85	25.67	<b>26.05</b>	23.92	22.93	23.96	<b>24.33</b>	22.67	21.52	22.66	<b>23.11</b>
R.R.Hood	28.65	28.10	28.73	<b>28.79</b>	27.31	26.75	27.41	<b>27.54</b>	26.25	25.73	26.45	<b>26.73</b>
Sailboats	28.68	27.93	28.68	<b>28.84</b>	27.10	26.34	27.14	<b>27.40</b>	25.84	25.19	26.00	<b>26.34</b>
Window	27.99	26.93	27.91	<b>28.28</b>	26.34	25.21	26.30	<b>26.60</b>	25.07	24.01	25.11	<b>25.41</b>
Baboon	21.81	21.58	21.98	<b>22.03</b>	20.83	20.57	<b>20.98</b>	20.94	20.21	19.97	<b>20.31</b>	20.25
Couple	26.40	25.41	26.23	<b>26.39</b>	<b>24.89</b>	23.90	24.81	24.82	23.81	22.88	<b>23.82</b>	23.73
F.boat	26.61	25.85	26.59	<b>26.62</b>	25.16	24.40	25.16	<b>25.19</b>	24.12	23.35	24.16	<b>24.20</b>
House	29.31	28.34	29.52	<b>30.20</b>	27.51	26.47	27.66	<b>28.50</b>	26.15	25.11	26.28	<b>27.12</b>
Peppers	26.59	25.71	26.63	<b>26.84</b>	24.97	23.83	24.90	<b>25.26</b>	23.67	22.40	23.57	<b>24.09</b>
straw	22.33	21.78	22.39	<b>22.51</b>	20.78	20.23	<b>20.88</b>	20.85	19.71	19.26	19.81	<b>19.68</b>
<b>Average</b>	26.54	25.82	26.58	<b>26.80</b>	25.02	24.22	25.05	<b>25.29</b>	23.87	23.07	23.93	<b>24.21</b>

where  $c_2 = c \cdot c_1$ .

It is not difficult to derive the solution to Eq. (15), which is basically a component wise soft-thresholding operation:

$$\vec{\alpha}_i = \vec{\mu}_i + \max\left(\left|\vec{\beta}_i - \vec{\mu}_i\right| - \frac{\sigma_n}{c_2 \cdot \sigma_i}, 0\right) \cdot \text{sgn}(\vec{\beta}_i - \vec{\mu}_i). \quad (16)$$

Then  $\mathbf{x}_i$  can be calculated by

$$\mathbf{x}_i = \Phi_i^T \vec{\alpha}_i. \quad (17)$$

After obtaining all the patches, we may get the full image  $\mathbf{x}$  by putting back the patches and averaging overlaps. Let  $E_i$  be the matrix extracting  $\mathbf{x}_i$  from  $\mathbf{x}$  at location  $i$ , i.e.  $\mathbf{x}_i = E_i \mathbf{x}$ , then the least-square solution writes [3]:

$$\mathbf{x} = \left(\sum_i E_i^T E_i\right)^{-1} \sum_i (E_i^T \mathbf{x}_i). \quad (18)$$

Besides, this paper borrows the wisdom of iterative regularization [8, 27] to update the observed noisy image in every iteration:

$$\tilde{\mathbf{y}} := \tilde{\mathbf{x}} + \rho(\mathbf{y} - \tilde{\mathbf{x}}), \quad (19)$$

and then regenerate  $\tilde{\mathbf{x}}$  using  $\tilde{\mathbf{y}}$  as the input image. Such technique has appeared in existing works, so detailed discussions are omitted here. Please refer to [8, 27] for further details. The proposed denoising procedure is summarized in Algorithm 1.

## 6. Experimental Results

This section evaluates the efficiency of the proposed method by comparing it with three state-of-art denoising schemes, including BM3D [2], LPG-PCA [13] and CSR

Table 2. SSIM Comparison (Unit: dB)

$\sigma_n$	10				20				30			
Schemes	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>
Airplane	<b>0.9355</b>	0.9309	0.9342	0.9343	<b>0.8844</b>	0.8734	0.8822	0.8808	<b>0.8531</b>	0.8373	0.8520	0.8429
Barbara	0.9515	0.9543	0.9540	<b>0.9552</b>	0.9036	0.9037	0.9067	<b>0.9100</b>	0.8499	0.8472	0.8543	<b>0.8585</b>
C.man	0.9283	0.9242	0.9239	<b>0.9298</b>	0.8684	0.8600	0.8644	<b>0.8751</b>	0.8256	0.8192	0.8284	<b>0.8338</b>
Lena	0.9422	0.9420	0.9430	<b>0.9448</b>	0.8907	0.8906	0.8954	<b>0.8965</b>	0.8462	0.8439	<b>0.8544</b>	0.8529
Monarch	0.9546	0.9554	0.9578	<b>0.9596</b>	0.9147	0.9154	0.9210	<b>0.9228</b>	0.8746	0.8747	<b>0.8859</b>	0.8865
R.R.Hood	0.8907	0.8879	0.8932	<b>0.8955</b>	0.8194	0.8137	<b>0.8250</b>	0.8239	0.7702	0.7677	<b>0.7826</b>	0.7745
Sailboats	<b>0.9287</b>	0.9285	0.9278	0.9284	0.8870	0.8831	<b>0.8857</b>	0.8850	0.8492	0.8428	<b>0.8535</b>	0.8447
Window	0.9543	0.9534	0.9558	<b>0.9563</b>	0.9126	0.9085	0.9151	<b>0.9170</b>	0.8725	0.8649	0.8797	<b>0.8803</b>
Baboon	0.8757	0.8627	0.8547	<b>0.8772</b>	0.7222	0.7091	0.7067	<b>0.7321</b>	0.6180	0.6039	0.6119	<b>0.6286</b>
Couple	0.8171	0.8453	<b>0.8454</b>	0.8369	0.7502	<b>0.7601</b>	0.7548	0.7564	<b>0.7865</b>	0.7616	0.7835	0.7844
F.boat	0.8867	0.8826	0.8854	<b>0.8889</b>	0.8197	0.8091	0.8184	<b>0.8198</b>	0.7682	0.7528	<b>0.7690</b>	0.7669
House	0.9150	0.9149	<b>0.9226</b>	0.9188	0.8657	0.8679	<b>0.8745</b>	0.8692	0.8374	0.8387	<b>0.8474</b>	0.8453
Peppers	0.9262	0.9225	0.9262	<b>0.9292</b>	0.8819	0.8768	0.8842	<b>0.8866</b>	0.8407	0.8373	<b>0.8494</b>	0.8491
straw	0.9573	0.9619	<b>0.9626</b>	0.9624	0.8948	0.8965	<b>0.9033</b>	0.9018	0.8233	0.8124	0.8295	<b>0.8314</b>
<b>Average</b>	0.9188	0.9190	0.9205	<b>0.9227</b>	0.8582	0.8549	0.8598	<b>0.8626</b>	0.8154	0.8075	<b>0.8201</b>	<b>0.8201</b>

$\sigma_n$	50				70				90			
Schemes	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>	<i>BM3D</i>	<i>L.PCA</i>	<i>CSR</i>	<i>Proposed</i>
Airplane	<b>0.8172</b>	0.7813	0.8132	0.7951	<b>0.7888</b>	0.7277	0.7849	0.7699	<b>0.7642</b>	0.6740	0.7597	0.7552
Barbara	0.7573	0.7412	<b>0.7615</b>	0.7594	0.6765	0.6557	<b>0.6870</b>	0.6803	0.6155	0.5870	0.6300	<b>0.6310</b>
C.man	0.7745	0.7586	<b>0.7803</b>	0.7799	0.7229	0.7003	0.7403	<b>0.7408</b>	0.6814	0.6419	0.7029	<b>0.7134</b>
Lena	0.7834	0.7667	0.7906	<b>0.7916</b>	0.7280	0.7018	0.7354	<b>0.7436</b>	0.6797	0.6430	0.6873	<b>0.7054</b>
Monarch	0.8142	0.7979	0.8205	<b>0.8229</b>	0.7582	0.7286	0.7661	<b>0.7716</b>	0.7084	0.6652	0.7166	<b>0.7298</b>
R.R.Hood	0.7180	0.7080	<b>0.7290</b>	0.7158	0.6769	0.6608	<b>0.6958</b>	0.6830	0.6436	0.6154	<b>0.6706</b>	0.6656
Sailboats	0.7744	0.7744	<b>0.8068</b>	0.7881	0.7563	0.7157	<b>0.7715</b>	0.7525	0.7167	0.6606	<b>0.7425</b>	0.7311
Window	0.8167	0.7871	<b>0.8216</b>	0.8199	0.7627	0.7169	0.7743	<b>0.7752</b>	0.7146	0.6517	0.7315	<b>0.7407</b>
Baboon	0.4727	0.4562	0.4819	<b>0.4934</b>	0.3881	0.3652	<b>0.3983</b>	0.3950	0.3358	0.3130	<b>0.3384</b>	0.3303
Couple	0.9072	0.8985	0.9060	<b>0.9077</b>	<b>0.8427</b>	0.8254	0.8401	0.8400	0.5812	0.5269	<b>0.5833</b>	0.5746
F.boat	0.6954	0.6689	<b>0.6959</b>	0.6879	0.6385	0.6070	<b>0.6446</b>	0.6342	0.5949	0.5551	<b>0.6055</b>	0.5973
House	0.8044	0.7848	<b>0.8126</b>	0.8116	0.7654	0.7309	0.7779	<b>0.7835</b>	0.7275	0.6776	0.7450	<b>0.7604</b>
Peppers	0.7859	0.7682	<b>0.7930</b>	0.7874	0.7342	0.7086	<b>0.7464</b>	0.7419	0.6891	0.6537	0.7047	<b>0.7062</b>
straw	0.6802	0.6480	0.6857	<b>0.6945</b>	0.5489	0.5000	<b>0.5607</b>	0.5525	0.4356	0.3837	<b>0.4399</b>	0.4196
<b>Average</b>	0.7572	0.7386	<b>0.7642</b>	0.7611	0.6991	0.6675	<b>0.7088</b>	0.7046	0.6349	0.5892	0.6470	<b>0.6472</b>



Figure 4. Test images for denoising experiments.

[8]. The proposed algorithm is implemented in Matlab 2010b, and the three anchor schemes are tested using executables or source codes released by the authors of the papers. Parameters used in the algorithm are empirically cho-

sen according to the noise levels so as to achieve relatively good performance. The detailed setting of three parameters are shown in Table 3, including patch size  $S$ , the number of similar patches in a group  $M$ , and the iterative regularization parameter  $\rho$  in (19). The denoising schemes are tested on 14 typical natural images with noise variance ranging from 10 to 100. The test images are displayed in Fig 4.

As shown in Table 1, the proposed scheme outperforms the other three methods in most cases in terms of peak signal to noise ratio (PSNR). When standard deviation of noise fluctuates from 10 to 90, the average PSNR of the proposed method is 0.46~1.14dB higher than LPG-PCA,

---

**Algorithm 1: Image Denoising by AST-NLS**

---

**Data:** The noisy image  $\mathbf{y}$   
**Result:** Denoised image  $\mathbf{x}$   
initialization:  $\mathbf{x} = \mathbf{y}, \tilde{\mathbf{y}} = \mathbf{y}$   
**while** *Outer stopping criteria unsatisfied* **do**  
    Update  $\tilde{\mathbf{y}}$  according to Eq. (19);  
    Update noise variance  $\sigma_n^2$ ;  
    **for**  $i = 1, 2, 3, \dots, N$  **do**  
        Calculate distances of nonlocal patches by Eq. (7);  
        Group similar patches of  $\tilde{\mathbf{y}}_i$  into  $Y_i$ ;  
        Apply PCA to patches in  $Y_i$ ;  
        Estimate expectations  $\vec{\mu}_i$  by calculating median of each band separately;  
        Estimate standard deviations  $\vec{\sigma}_i$  by Eq. (12) and (11);  
        Operate soft-thresholding according to Eq. (16);  
        Calculate  $\mathbf{x}_i$  by Eq. (17);  
    **end**  
    Aggregate estimated patches into  $\mathbf{x}$  by Eq. (18);  
**end**

---

Noise level	$\rho$	$M$	$S$
$\sigma_n \leq 10$	0.11	75	6
$10 < \sigma_n \leq 20$	0.11	80	7
$20 < \sigma_n \leq 40$	0.12	100	7
$40 < \sigma_n \leq 60$	0.12	130	8
$60 < \sigma_n \leq 80$	0.12	150	9
$\sigma_n > 80$	0.12	150	10

Table 3. Parameter settings.

0.26~0.39dB higher than BM3D and 0.11~0.28dB higher than CSR. The structural similarity (SSIM) [28] of AST-NLS also outperforms BM3D and LPG-PCA, and is highly competitive against CSR, as displayed in Table 2.

Our ultimate goal is to achieve better perceptual quality since in most cases images are meant to be viewed by people’s eyes. Fig. 5, Fig. 6, Fig. 7 and Fig. 8 demonstrate visual improvement of the proposed scheme at different noise levels. It is evident that the images denoised by the proposed scheme exhibit much less noise and artifacts (*e.g.* the face of the man and the background in Fig. 5), and better preserves details and textures (*e.g.* the tentacles of monarch in Fig. 7). The images denoised by LPG-PCA exhibit slightly less noise and artifacts than BM3D and CSR but tend to be blurred. Comparing with the three anchor schemes, the output of the proposed scheme is obviously more clean, sharp and visually pleasant.

## 7. Conclusions

This paper aims to develop an effective image denoising scheme based on adaptive distribution estimation in transform domain. The regularization attempts to approxi-

mate the actual distribution of transform coefficients. Since the statistical characteristics may vary significantly, image patches at different locations are processed separately, and coefficients in different bands are treated distinctively. We use a set of nonlocal similar patches as samples to adaptively estimate the distribution parameters, including expectation and variance. Experimental results demonstrate that the proposed denoising scheme outperforms some state-of-art methods like BM3D and CSR, in terms of both objective measurements and perceptual quality.

## References

- [1] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *IEEE International Conference on Computer Vision*, Jan 1998, pp. 839–846.
- [2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.
- [3] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.
- [4] A. Foi, V. Katkovnik, and K. Egiazarian, “Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images,” *IEEE Transactions on Image Processing*, vol. 16, no. 5, pp. 1395–1411, May 2007.
- [5] D.L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [6] D.D. Muresan and T.W. Parks, “Adaptive principal components and image denoising,” in *IEEE International Conference on Image Processing*, Sept 2003, vol. 1, pp. 101–104.
- [7] A. Buades, B. Coll, and J.M. Morel, “A non-local algorithm for image denoising,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2005.
- [8] W. Dong, X. Li, D. Zhang, and G. Shi, “Sparsity-based image denoising via dictionary learning and structural clustering,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 457–464.
- [9] L.I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.
- [10] L. Sendur and I.W. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744–2756, 2002.
- [11] A. Beck and M. Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.
- [12] P. Chatterjee and P. Milanfar, “Clustering-based denoising with locally learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1438–1451, July 2009.



Figure 5. Perceptual comparison for *Cameraman*. From left to right: (a) Noisy image ( $\sigma_n = 30$ ); (b) BM3D (PSNR = 28.27dB); (c) LPG-PCA (PSNR = 27.83); (d) CSR (PSNR = 28.54dB); (e) Proposed (PSNR = 28.63dB).



Figure 6. Perceptual comparison for *Window*. From left to right: (a) Noisy image ( $\sigma_n = 30$ ); (b) BM3D (PSNR = 30.49dB); (c) LPG-PCA (PSNR = 29.76); (d) CSR (PSNR = 30.63dB); (e) Proposed (PSNR = 30.92dB).

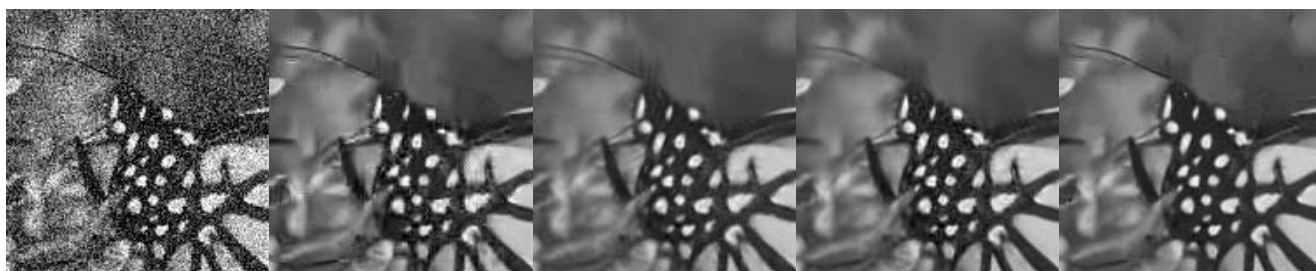


Figure 7. Perceptual comparison for *Monarch*. From left to right: (a) Noisy image ( $\sigma_n = 40$ ); (b) BM3D (PSNR = 25.77dB); (c) LPG-PCA (PSNR = 26.14); (d) CSR (PSNR = 26.81dB); (e) Proposed (PSNR = 27.17dB).

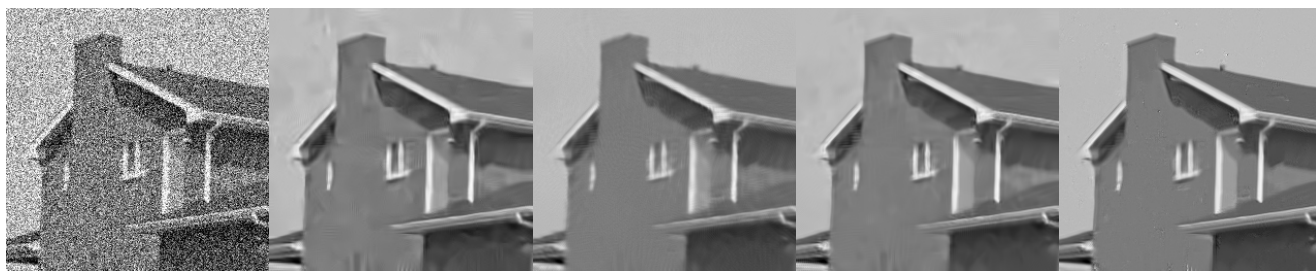


Figure 8. Perceptual comparison for *House*. From left to right: (a) Noisy image ( $\sigma_n = 50$ ); (b) BM3D (PSNR = 29.31dB); (c) LPG-PCA (PSNR = 28.34); (d) CSR (PSNR = 29.52dB); (e) Proposed (PSNR = 30.20dB).

[13] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, no. 4, pp. 1531–1549, 2010.

[14] S.G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[15] G. Yu and G. Sapiro, "DCT image denoising: a simple and effective image denoising algorithm," *Image Processing On Line*, vol. 1, 2011, <http://dx.doi.org/10.5201/ipol.2011.ys-dct>.

[16] J. Zhang, R. Xiong, C. Zhao, S. Ma, and D. Zhao, "Exploiting image local and nonlocal consistency for mixed gaussian-impulse noise removal," in *IEEE International Conference on Multimedia and Expo*, July 2012, pp. 592–



- [17] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1635–1649, April 2012.
- [18] M. Lebrun, A. Buades, and J.M. Morel, "A nonlocal bayesian image denoising algorithm," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1665–1688, 2013.
- [19] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, vol. 2, pp. 2862–2869.
- [20] H. Liu, R. Xiong, S. Ma, X. Fan, and W. Gao, "Non-local extension of total variation regularization for image restoration," in *IEEE International Symposium on Circuits and Systems*, Jun. 2014, pp. 1102–1105.
- [21] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 915–928, Jun. 2014.
- [22] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.
- [23] R. Rubinstein, A.M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.
- [24] H. Liu, R. Xiong, S. Ma, X. Fan, and W. Gao, "Gradient based image transmission and reconstruction using non-local gradient sparsity regularization," in *IEEE International Conference on Multimedia and Expo*, Jul. 2014, pp. 1–6.
- [25] H. Liu, R. Xiong, S. Ma, X. Fan, and W. Gao, "Gradient based image/video softcast with grouped-patch collaborative reconstruction," in *IEEE Visual Communications and Image Processing Conference*, Dec. 2014, pp. 141–144.
- [26] J. Zhang, R. Xiong, S. Ma, and D. Zhao, "High-quality image restoration from partial random samples in spatial domain," in *IEEE Visual Communications and Image Processing Conference*, Nov. 2011, pp. 1–4.
- [27] S. Osher, M. Burger, D. Goldfarb, J. Xu, , and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 460–489, 2005.
- [28] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.