NEW BOUNDS ON IMAGE DENOISING: VIEWPOINT OF SPARSE REPRESENTATION AND NON-LOCAL AVERAGING

Jianzhou Feng[1,2]*, Li Song*[1,2]*, Xiaoming Huo*[3]*, Xiaokang Yang*[1,2] *and Wenjun Zhang*[1,2]

[1] Shanghai Digital Media Processing and Transmission Key Lab
[2]Shanghai Jiao Tong University, Shanghai, China 200240
E-mail: {bravefjz, song_li (corresponding author), xkyang, zhangwenjun}@sjtu.edu.cn
[3] Georgia Institute of Technology, Atlanta, GA, U.S.A 30332
E-mail: xiaoming@isye.gatech.edu

ABSTRACT

Image denoising plays a fundamental role in many image processing applications. Utilizing sparse representation and non-local averaging together is such a successful framework that leads to considerable progress in denoising. Almost all the newly proposed denoising algorithms are built base on it, different in detailed implementation, and the denoising performance seems converging. What is the denoising bound of this framework turns into a key question. In this paper, we assume all the possible algorithms under the framework can be approximated by a fixed two steps denoising process with different parameters. Step one cluster geometric similar image patches into groups so that patches within each group could be sparse represented under the basis of the group. Step two use the atoms of the group basis and radiometric similar patches of each patch for non-local averaging. The parameters of the process are the cluster number, the atoms and the number of radiometric similar patches for estimating each patch. Finally, the bound is derived as the minimum denoising error of all the possible parameters. Comparing with previous bounds, the new one is image specific and more practical. Experiment results show that there still exists room to improve the denoising performance for natural images.

***Index Terms*—** Image specific bound, affine bias model, Fisher information matrix

1. INTRODUCTION

Image denoising plays a fundamental role in many image processing applications. In the past 50 years, the denoising problem has been analyzed under different prior assumptions. There are Total Variation based methods, transform domain filtering based methods, and Markov Random Field based methods such as the Fields of Experts. Nowadays, state-of-the-art algorithms, BM3D [1], learned simultaneous sparse coding [2] and clustering-based sparse representation [3] have lead to a significant progress on the denoising results. However, all these algorithms produce quite comparable results and they share the same framework, which utilizes sparse representation and non-local averaging. This makes us to think: Have we reached the theoretical limit of the denoising performance?

There are several works aims to study the limits of image denoising. The most recent two literature are given by Levin *et al.* [4] and Chatterjee *et al.* [5, 6]. In [4], a generic image prior is approximated by a huge set of 10^{10} natural image patches and a lower bound on the optimal Bayesian Minimum Mean Square Error is provided. However, image prior of a specific image may be quite different from the generic one and all the state-of-the-art denoising algorithms take advantage of adapting to the observed image. This makes the comparison between the bound using the generic image prior and the denoising results of nowadays algorithms unfair. In [5, 6], the authors derived a denoising bounds that do account for specific image statistics. They factorized the image patches of a specific image into a small number of clusters containing geometric similar patches and estimated the bounds via an affine bias model in each cluster. However, the assumption in their work, radiometric similar patches are exactly equal to the reference patch, is too strong and large model error is introduced between the real denoising performance and the theoretical bounds.

In this paper, we derive a more practical bound under the commonly used denoising framework, utilizing sparse representation and non-local averaging. We simulate the framework and derive the bound as:

1. Geometric similar image patches are clustered into K groups, *Principle Component Analysis* (PCA) is applied to the patches so that patches could be sparse represented by the PCA basis $\mathbf{U}_k$ of group k.
2. For each patch $\mathbf{z}_i$ in group k, N_i radiometric similar patches and matrix $\mathbf{U}_{k,i}$, composed by a few atoms of $\mathbf{U}_k$, are used for non-local averaging under an affine bias model.
3. Calculate the *Mean Squared Error* (MSE) of the de-

noising result under all possible model parameter set $(K, \{N_i, \mathbf{U}_{k,i}\}_i)$'s. The bound is derived as the minimum one.

Our work borrows the idea in [5, 6], such as clustering the geometric similar image patches and denoising the image patch using the affine bias model, but with three main differences:

- The geometric similar image patches are clustered for utilizing sparse representation.
- Radiometric similar patches are used in affine bias model instead of the geometric similar ones for utilizing non-local averaging.
- The proposed bound is the real denoising result under the optimized model parameter set, which should be more practical than the theoretical one with a too strong assumption.

Experiment results show that there still exists room to improve the denoising performance for natural images.

The remainder of the paper is organized as follows: Section 2 analyzes the existing denoising bounds proposed in [4, 5, 6]. Section 3 presents the model for deriving the new bound and compares the bound with the denoising results of BM3D on several standard images. Section 4 concludes the paper.

2. ANALYZING THE EXISTING BOUNDS

Most recent denoising algorithms [1, 2, 3, 7, 8] and models for deriving denoising bounds [4, 5, 6] are all based on image patches. So, we first present the patch-based image model as

$$\mathbf{y}_i = \mathbf{z}_i + \eta_i, \tag{1}$$

where $\mathbf{z}_i$ is the vectorized form of the actual patch intensity, $\mathbf{y}_i$ corresponds to the observed patch and η_i is a vectorized noise patch. We assume the noise to be white gaussian and the patches are nonoverlapping which makes η_i's uncorrelated. In this section, we shall analyze and show the limitations of the existing bounds.

2.1. The bound in [4]

The bound in [4] is derived using a generic image prior, whereas an image specific prior can lower the MSE of the estimation under the Bayasian estimation framework. Here is a toy example, which compares the performance of using generic image prior with the image specific one.

As shown in Fig. 1, there are three 64×64 images, the pixel intensity at position (i, j) of the images are $i + 96$, 128, $j + 96$ respectively. The image specific prior $p_k(\mathbf{z})$, $k = 1, 2, 3$, of the k^{th} image is described by all the 8×8 patches from image k and the generic prior as $p(\mathbf{z}) = (p_1(\mathbf{z}) + p_2(\mathbf{z}) + p_3(\mathbf{z}))/3$. We compute the Bayesian estimate of noisy patch $\mathbf{y}_i$'s from the three images under two different priors:

(a) (b) (c)

Fig. 1. Three testing images

Table 1. The SNR gap between image specific prior and the generic prior

σ	5	15	25	50
SNR gap(dB)	0.9	2.4	1.9	1.7

- Generic prior estimate: $\hat{\mathbf{z}}_i = \int_z p(\mathbf{y}_i|z)p(z)dz$.
- Specific prior estimate: $\hat{\mathbf{z}}_i = \int_z p(\mathbf{y}_i|z)p_k(z)dz$, k corresponds to which image $\mathbf{y}_i$ comes from.

The SNR gap is defined as

$$10\log_{10}\left(\frac{\text{BMMSE}_\text{g}}{\text{BMMSE}_\text{s}}\right), \tag{2}$$

where BMMSE_g stands for the *Bayesian Minimum Mean Square Error* (BMMSE) using $p(\mathbf{z})$ and BMMSE_s stands for using $p_k(\mathbf{z})$'s. As shown in Table. 1, the averaged gap is more than 1.5dB which is significant in image denoising.

Most nowadays denoising algorithms have already used the image specific information from the observed image. So the 0.1 dB margin shown in [4] actually underestimate the margin between the image specific bound and the denoising performance of the state-of-the-art denoising algorithms.

2.2. The bound in [5, 6]

Another thread of deriving the bound is by Chatterjee *et al.* [5, 6]. They first clustered the geometric similar image patches together from the observed image. All the image patches $\mathbf{z}_i$'s in the same cluster are assumed to be sampled values of a random variable $\mathbf{z}$. For each estimator $\hat{\mathbf{z}}$, the bias $b(\mathbf{z})$ is modeled as

$$\mathbf{b}(\mathbf{z}) = E[\hat{\mathbf{z}}|\mathbf{z}] - \mathbf{z} = \mathbf{M}\mathbf{z} + \mathbf{u}. \tag{3}$$

It is an affine bias model whose parameters are matrix $\mathbf{M}$ and vector $\mathbf{u}$. Under this model, the theoretical bound for patch $\mathbf{z}_i$ in cluster k is derived as

$$E\left[\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2\right] \geq \text{Tr}\left\{\left(\mathbf{J}_i + \text{cov}_k^{-1}(\mathbf{z})\right)^{-1}\right\}, \tag{4}$$

where $\mathrm{cov}_k(\mathbf{z})$ is the covariance matrix of patches in cluster k and $\mathbf{J}_i$ is the conditional Fisher information matrix (FIM). Then the authors define the radiometric similar patches $\mathbf{z}_j$'s to $\mathbf{z}_i$ that satisfy

$$\|\mathbf{z}_j - \mathbf{z}_i\|^2 \leq \gamma, \tag{5}$$

γ is a small threshold value. They further assume all $\mathbf{z}_j$'s are exactly equal to $\mathbf{z}_i$ so that

$$\mathbf{J}_i = -E\left[\frac{\partial^2 \ln p(\mathbf{y}|\mathbf{z})}{\partial \mathbf{z}_i \partial \mathbf{z}_i^T}\right] = N_i \frac{\mathbf{I}}{\sigma^2}, \tag{6}$$

where N_i is the number of $\mathbf{z}_j$'s. In fact, for deriving (6), the assumption "N identical patches are considered", as shown in [5], can be relaxed to

$$\frac{1}{N_i}\sum \mathbf{z}_j = \mathbf{z}_i. \tag{7}$$

We shall show (7) is too strong an assumption which makes the bound unrealistic to achieve. Let's analyze the estimator

$$\hat{\mathbf{z}}_i = (\mathbf{I} + \mathbf{M}_i)\frac{\sum \mathbf{y}_j}{N_i} + \mathbf{u}_i, \tag{8}$$

where $\mathbf{M}_i$ and $\mathbf{u}_i$ are the optimal parameter that makes

$$\begin{aligned}&\int_z \left[\mathrm{Tr}\left\{(\mathbf{I}+\mathbf{M}_i)\mathbf{J}_i^{-1}(\mathbf{I}+\mathbf{M}_i)^T\right\} + \|b(\mathbf{z})\|^2\right] p_k(\mathbf{z})d\mathbf{z} \\ &= \mathrm{Tr}\left\{\left(\mathbf{J}_i + \mathrm{cov}_k^{-1}(\mathbf{z})\right)^{-1}\right\}.\end{aligned} \tag{9}$$

as shown in [5]. If the assumption is true, then the estimator can achieve the theoretical bound. Firstly, it belongs to the affine model because

$$\mathbf{b}(\mathbf{z}_i) = E[\hat{\mathbf{z}}_i|\mathbf{z}_i] - \mathbf{z}_i = \mathbf{M}_i\mathbf{z}_i + \mathbf{u}_i. \tag{10}$$

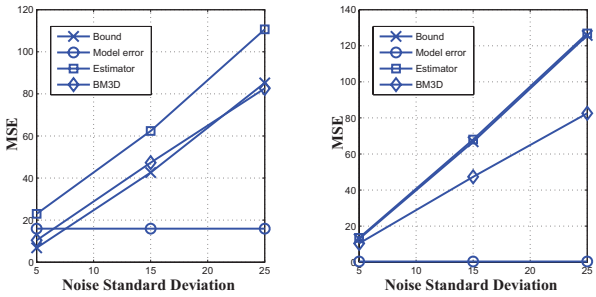
Secondly, the expected MSE

$$\begin{aligned}&E\left[\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2\right] = \int_{z_i} E\left[\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2|\mathbf{z}_i\right] p_k(\mathbf{z}_i)d\mathbf{z}_i \\ &= \int_{z_i} \left(E\left[\|\hat{\mathbf{z}}_i - E[\hat{\mathbf{z}}_i|\mathbf{z}_i]\|^2|\mathbf{z}_i\right] + \|b(\mathbf{z}_i)\|^2\right)p_k(\mathbf{z}_i)d\mathbf{z}_i \\ &= \mathrm{Tr}\left\{(\mathbf{I}+\mathbf{M}_i)\mathbf{J}_i^{-1}(\mathbf{I}+\mathbf{M}_i)^T\right\} + \int_{z_i}\|b(\mathbf{z}_i)\|^2 p_k(\mathbf{z}_i)d\mathbf{z}_i \\ &= \mathrm{Tr}\left\{\left(\mathbf{J}_i + \mathrm{cov}^{-1}(\mathbf{z})\right)^{-1}\right\}.\end{aligned} \tag{11}$$

is exactly the bound. So the denoising MSE of applying such an estimator to the degraded image should be very close to the theoretical bound if the assumption is realistic.

In Fig. 2, using the software released by the authors, we show the theoretical bound, the MSE of the estimator and the model error

$$E\left[\left\|\mathbf{z}_i - \frac{1}{N_i}\sum \mathbf{z}_j\right\|^2\right] \tag{12}$$



(a) $\gamma = (\frac{5\times255}{100})^2 \times n$ (b) $\gamma = (\frac{1\times255}{100})^2 \times n$

Fig. 2. MSE under different noise level and γ of Parrot

under different threshold value γ, where n is the dimension of $\mathbf{z}$. For comparison, we also present the curve of denoising result by applying BM3D [1] to the degraded images, which is shown repeatedly under different γ. As expected, the gap between the theoretical bound and the presented estimator is very close to the model error. It turns out that: When γ is large, though the theoretical bound is low, the large model error makes it unrealistic to achieve. When γ is small, even the bound itself is worse than BM3D.

3. THE NEW BOUND AND EXPERIMENTS

3.1. The new bound

In this paper, we aim to derive a practical image denoising bound under the framework which utilize both sparse representation and non-local averaging. Since nowadays denoising algorithms [1, 2, 3, 7, 8] are mainly under this framework, just different in detailed techniques, such a bound is helpful for answering the following two questions

- Has the best performance of the framework been reached by nowadays algorithms?
- Is it valuable for designing a new work using new techniques under the same framework?

We simulate the framework into a simplified two step denoising process:

1. Geometric similar image patches are clustered into K groups. The PCA basis $\mathbf{U}_k$ is derived from noise patch $\mathbf{y}_i$'s in cluster k. $\mathbf{U}_k$ capture the similar local structure and singularity of $\mathbf{z}_i$'s, which makes it a better basis for representation $\mathbf{z}_i$'s than a predefined basis, such as DCT. From the view of sparse representation, we are using a structured dictionary composed by K orthogonal basis $\mathbf{U}_k$'s and atoms from different $\mathbf{U}_k$'s are not allowed for representing one patch simultaneously. The structured dictionary has been used in [9] with good performance in denoising.

2. For each patch $\mathbf{z}_i$, we find its N_i radiometric similar patches $\mathbf{z}_j$'s, then estimate $\mathbf{z}_i$ using $\mathbf{y}_j$'s and $\mathbf{U}_k$, k as the index of group $\mathbf{z}_i$ belongs to. The radiometric similar patches are a supplementary to sparse representation. Sparse representation can only remove the noise outside the represented subspace, whereas natural patches owns the non-local similarity, which means they are always locate close to several centers in the subspace, so that averaging on $\mathbf{y}_j$'s can help to shrink the noise within the subspace.

In step 1, the Locally Adaptive Regression Kernels (LARK) of each patch is used as the geometric feature and K-means is used for clustering as in [5, 6]. In step 2, $\mathbf{z}_i$ is estimated under the affine model. The affine model could simulate any denoising algorithm for radiometric similar patches. Because applying Taylor expansion to the bias function $\mathbf{b}(\mathbf{z})$ at $\mathbf{z}_j$, we have

$$\begin{aligned}\mathbf{b}(\mathbf{z}_j) &= \mathbf{b}(\mathbf{z}_i) + \frac{\partial \mathbf{b}(\mathbf{z}_i)}{\partial \mathbf{z}}(\mathbf{z}_j - \mathbf{z}_i) + O(\|\mathbf{z}_j - \mathbf{z}_i\|^2) \\ &\approx \frac{\partial \mathbf{b}(\mathbf{z}_i)}{\partial \mathbf{z}}\mathbf{z}_j + (\mathbf{b}(\mathbf{z}_i) - \frac{\partial \mathbf{b}(\mathbf{z}_i)}{\partial \mathbf{z}}\mathbf{z}_i) \\ &= \mathbf{M}(\mathbf{z}_i)\mathbf{z}_j + \mathbf{u}(\mathbf{z}_i).\end{aligned} \tag{13}$$

Using a set of columns of $\mathbf{U}_k$, denote as $\mathbf{U}_{k,i}$, and $\mathbf{y}_j$'s, the mean vector of $\mathbf{z}_j$'s is estimated as

$$\mathbf{u}_i = \frac{1}{N_i}\sum_j U_{k,i}U_{k,i}^T\mathbf{y}_j \tag{14}$$

and the Covariance Matrix as

$$\mathrm{Cov}_i = \mathbf{U}_{k,i}[\mathrm{Cov}(\{\mathbf{U}_{k,i}^T\mathbf{y}_j\}) - \sigma^2 I]_+\mathbf{U}_{k,i}^T, \tag{15}$$

where $[\mathbf{X}]_+$ denotes a matrix with the negative eigenvalues of $\mathbf{X}$ replaced by 0. Finally, the estimator $\hat{\mathbf{z}}_i$ under the affine model is

$$\hat{\mathbf{z}}_i = \mathbf{u}_i + \mathrm{Cov}_i(\mathrm{Cov}_i + \sigma^2 I)^{-1}(\mathbf{y}_i - \mathbf{u}_i). \tag{16}$$

All the denoising algorithms under the framework can be seem as using different parameter set $(K, \{N_i, \mathbf{U}_{k,i}\})$. So we derive the bound as

$$\min_{K,\{N_i,\mathbf{U}_{k,i}\}} \frac{1}{nN}\sum_i \|\hat{\mathbf{z}}_i - \mathbf{z}_i\|_2^2, \tag{17}$$

where n is the dimension of $\mathbf{z}_i$ and N is the number of $\mathbf{z}_i$'s. If a denoising algorithm can estimate the parameter set close to the optimum one, the denoising performance could also be close to the bound. In comparison to the previous bounds, the new bound is more practical because it is obtained by denoising the noisy patches instead of the theoretical ones as in [5, 6] and the only assumption we make is the noisy free patches are available for radiometric similar patches matching. In state-of-the-art denoising algorithms, the final block matching is always performed on a prefiltered estimate other than the noisy one. As we don't constrain the prefiltering algorithm and the noise level is not high, we assume the prefiltered estimate to be close to the clean original image. We will validate this assumption in section 3.2.

3.2. Experiment results

In the experiment, we derive the bound on several standard images, as shown in Fig. 3. In order to validate using clean patches for block matching, we assume the BM3D prefiltered image is used for finding similar patches so that also derive a sub-optimal bound. Finally, these two bounds are compared with the results of BM3D [1]. As in [5], the patch size is fixed to be 11×11 so that $n = 121$, the possible parameter set is constrained to be $K \in \{5, 10, 15, 20, 25\}$, $N_i \in [1, 20]$ and $\mathbf{U}_{k,i}$ could use any set of atoms in $\mathbf{U}_k$. The experiment results listed in Table 2 shows the gap between the bound and the sub-optimal one is neglectable, which validate the assumption, and there still exists room to improve the denoising performance for natural images. Using the two step denoising process, if an algorithm could estimate the optimal parameter set accurately, then better denoising result could be achieved. As there is no ground truth of the real bound, comparison between different bounds are mainly based on the assumptions. Compare with the bound in [4], the averaged PSNR improvements under $\sigma = 25$ is about 1dB, much larger than 0.1dB in [4]. The higher PSNR gap probably benefits from using the image specific information. Compare with the bound in [5], the bound of house under $\sigma = 25$ is 14.82, much less than our bound 24.26 and the BM3D result 33.62. The bound in [5] probably gives overoptimistic results. Because using the same γ, flat image like *house* may lead to too large radiometric similar patch number N, actually not so many identical patches, so that too small bound. The bound in [6] is almost the same as [5] so there is no need for further comparison.

4. CONCLUSION

In this paper, we present a new image specific bound for natural image denoising. The new bound is derived under the framework utilizing sparse representation and non-local averaging. Experiment results show that there still exists room to improve in image denoising area. The future work includes exploring the relationship between the optimal model parameter set and the degraded image, then designing a denoising algorithm to approach the bound.

5. ACKNOWLEDGMENT

This work was supported by the 973 Program (2010CB731401, 2010CB731406), the china MIIT Program (2010ZX03004-003), the national 863 Program (2012AA011703), the NSF grants (60932006, 60902020,

parrot peppers

house cameraman

Fig. 3. Four standard images

and 61102098), STCSM (12DZ2272600) and the 111 project.

6. REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, 2007.

[2] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. ICCV*, 2009.

[3] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. CVPR*, 2011.

[4] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *Proc. CVPR*, 2011.

[5] P.Chatterjee and P.Milanfar, "Is denoising dead?," *IEEE Trans. Image Process.*, 2010.

[6] P.Chatterjee and P.Milanfar, "Practical bounds on image denoising: From estimation to information," *IEEE Trans. Image Process.*, 2011.

[7] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, 2006.

Table 2. The bound, the sub-optimal bound (Bound*) and BM3D MSE under different noise level

	parrot	peppers	house	cameraman
BM3D	10.63	10.02	6.76	9.64
Bound	9.20	8.93	5.16	8.64
Bound*	9.26	9.01	5.38	8.77

$\sigma = 5$

	parrot	peppers	house	cameraman
BM3D	47.37	34.91	20.83	41.84
Bound	37.89	31.03	14.94	36.02
Bound*	38.51	31.50	16.56	36.68

$\sigma = 15$

	parrot	peppers	house	cameraman
BM3D	83.10	62.65	33.62	73.76
Bound	65.18	52.04	24.26	61.38
Bound*	66.41	53.29	27.50	63.07

$\sigma = 25$

[8] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multisc. Model. Simulat.*, 2005.

[9] G. Yu, G. Sapiro, and S. Mallat, "Image modeling and enhancement via structured sparse model selection," in *Proc. IEEE Int. Conf. Image Process.*, Sept. 2010, pp. 1641–1644.