LEARNING DENOISING BOUNDS FOR NOISY IMAGES

Priyam Chatterjee and Peyman Milanfar

Department of Electrical Engineering
University of California, Santa Cruz, CA 95064, USA
{priyam, milanfar}@soe.ucsc.edu

ABSTRACT

In [1], we derived an expression for the fundamental limit to image denoising assuming that the noise-free image is available. In this paper, we propose an *estimator* for the bound on the mean squared error given only the noisy image and noise characteristics. To do this, we make use of an assortment of independently collected noise-free images from which prior information about the noisy image is learned. We show that even for reasonably low input signal-to-noise levels, our method can predict the denoising bound with accuracy.

Index Terms— Image denoising, estimation, Bayesian Cramér-Rao lower bound, mean squared error.

1. INTRODUCTION

Image denoising has been a well studied problem in the image processing community. Recently published methods perform denoising quite well, even in the presence of high levels of noise. Camera manufacturers rely on these methods to attenuate the effects of noise, especially for images captured in unfavorable lighting conditions. However, not much effort has been made to study the fundamental performance limits of image denoising. Recently, Treibitz *et al.* [2] studied the recovery limits for particular objects or regions in an image under pointwise degradation. Voloshynovskiy *et al.* [3] briefly analyzed the performance of MAP estimators for the denoising problem. However, neither of these works study the general image denoising problem. In [1], we developed bounds on the mean squared error (MSE) of patch-based denoising methods. This enabled us to identify how well the current state-of-the-art methods performs when compared to the theoretical limits of performance. The bounds there were computed assuming knowledge of the noise-free image.

In this work, we propose a method of estimating the bounds given only a *noisy* observation. If built into a camera, such a method will make it possible to predict if a captured image, once denoised, will be of acceptable quality. This can be useful to photographers who can then tune camera parameters accordingly, manually or automatically. In the absence of the ground truth, we make use of a collection of noise-free images that is processed offline to infer information about the noisy image. We also use a modified patch similarity measure to determine patch redundancy in the noisy image. The bound estimation process is detailed in Sec. 3. However, we first provide a brief overview of our bounds formulation in the next section. In Sec. 4, we validate our method proposed in this paper through various experiments that amply illustrate that the bounds can be accurately estimated from images corrupted by varying levels of noise. We conclude in Sec. 5 where we identify directions for future research.

This work was supported in part by the U.S. Air Force under Grant F49620-03-1-0387.

(a) (b) (c) (d) (e)

Fig. 1. Example of geometric clustering: (a) Noise-free house image, (b)-(e) few clusters based on geometric structure of patches.

2. LOWER BOUND ON THE MSE

In [1], we studied the fundamental limits of image denoising where the problem is to estimate the original image patches $\mathbf{z}_i$ from their noisy observations

$$\mathbf{y}_i = \mathbf{z}_i + \boldsymbol{\eta}_i, \qquad i = 1, 2, \ldots, M \tag{1}$$

where M is the number of patches in the image, and $\boldsymbol{\eta}_i$ denotes a noise patch. An expression for the Bayesian Cramér-Rao lower bound on the MSE for any given image was derived by an independent analysis of various geometric structures in the image. The motivation for such an approach was that, under the assumption of the noise being independent of pixel intensity, denoising difficulty is proportional to the level of detail present in the underlying image patch. Thus, the (noise-free) image patches were segmented into clusters of geometrically similar patches (Fig. 1), irrespective of their pixel intensities, using normalized steering kernel features developed for denoising in [4, 5]. This allowed us to model patches $\mathbf{z}_i$ in any given cluster (Ω_k) as realizations of some random variable $\mathbf{z}$ sampled from some (unknown) probability density function

(pdf) $p_k(\mathbf{z})$. The MSE bound for each patch within a cluster was then derived as

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

where $\mathbf{J}_i$ is the Fisher information matrix (FIM) and $\mathbf{C}_{\mathbf{z}}$ is the covariance of $\mathbf{z}$ from pdf $p_k(\mathbf{z})$. $\mathbf{C}_{\mathbf{z}}$ captures the cluster complexity in terms of variation between member patches while the FIM depends on the noise characteristics and the number (N_i) of similar patches that exist for each patch ($\mathbf{z}_i$). For zero mean Gaussian noise $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, the FIM takes the form

$$\mathbf{J}_i = N_i \frac{\mathbf{I}}{\sigma^2} \tag{3}$$

where $\mathbf{I}$ denotes the identity matrix. N_i is determined patch-wise for each $\mathbf{z}_i$ by searching for the number of patches $\mathbf{z}_j$ in the entire image that satisfy the condition

$$\mathbf{z}_j = \mathbf{z}_i + \boldsymbol{\varepsilon}_{ij} \quad \text{such that } \|\boldsymbol{\varepsilon}_{ij}\|^2 \leqslant \gamma \tag{4}$$

where γ is a patch size dependent threshold. The bound for a cluster Ω_k can then be computed as

$$E\left[\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2\right]_{\Omega_k} \geqslant \frac{1}{M_k}\sum_{i=1}^{M_k} \mathrm{Tr}\left[\left(\mathbf{J}_i + \mathbf{C}_{\mathbf{z}}^{-1}\right)^{-1}\right] \tag{5}$$

for all $\mathbf{z}_i \in \Omega_k$ with cluster cardinality M_k. Note that the expression for the bound does not require knowledge of entire $p_k(\mathbf{z})$ but just its first and second order moments, which are estimated using a bootstrapping mechanism [6]. However, in [1], the moments and the N_i values were estimated from the noise-free image that was assumed to be available. In practice, one may need to estimate the bound given only the noisy image and some knowledge of the noise characteristics. We propose a way of *estimating* the bound for such cases in the next section.

3. LOWER BOUND FROM NOISY IMAGE

As previously mentioned, the bound estimation process in [1] relies on information accurately gathered from the noise-free image. However, given a noisy image, the estimation of the covariance matrix in each cluster and the performance of clustering are affected by the presence of noise. Also dependent on the presence of the noise-free image is the estimation of the N_i values for each patch in the image. In our present approach, we estimate the required information, namely $\mathbf{C}_{\mathbf{z}}$ and N_i values, using an approach graphically outlined in Fig. 2. We describe each step below.

3.1. Estimating the Covariance Matrix

In the presence of noise, the image patches can no longer be clustered without noticeable clustering error. Such errors automatically reduce the accuracy of the covariance estimate. Moreover, even if perfect clustering was possible, noise would adversely affect the estimation of the second moment from the noisy image samples.

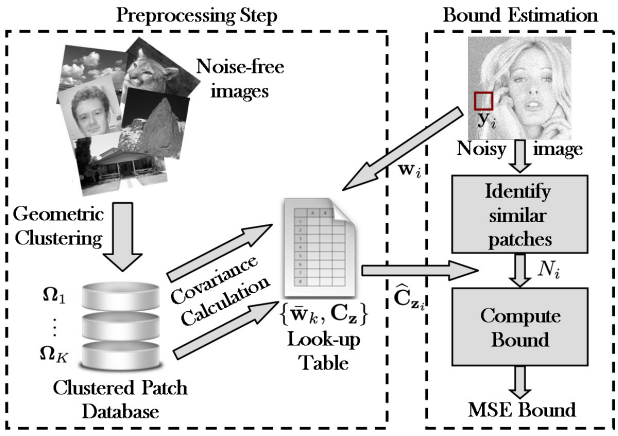


Fig. 2. Block diagram of our proposed bound estimation method.

To circumvent such adverse effects of noise, we make use of a collection of noise-free images (not containing the target underlying image) that is processed offline. Extracting patches from such images, we compute the steering kernels for each noise-free patch using the method outlined in [4, 5]. These steering kernels capture the underlying patch structure and are used as features for clustering the noise-free patches across all images in the database. We use the K-Means method [7] to group the noise-free patches into a large number (K) of clusters (denoted by Ω_k). Along with the clustering, K-Means also estimates the cluster centers ($\bar{\mathbf{w}}_k$). Once the clusters are formed, we can expect patches of similar geometric structure to be grouped together (Fig. 1), irrespective of the (noise-free) image that they belong to. A bootstrapping mechanism [6] is then used to estimate the covariance matrix for each cluster. Once this is done, we only retain the cluster centers and the associated $\mathbf{C}_{\mathbf{z}}$ matrices to form a look-up table. This forms the preprocessing step of our method, as shown in Fig. 2.

Now, given any noisy image, we obtain an estimate of the covariance matrix for each *noisy* patch by identifying the cluster of *noise-free* patches that it is most likely to belong to. This is done by computing the steering kernel features $\mathbf{w}_i$ for each patch in the noisy image and associating with each noisy patch some cluster Ω_κ with center $\bar{\mathbf{w}}_\kappa$ from the precomputed look-up table such that

$$\kappa = \arg\min_k \|\mathbf{w}_i - \bar{\mathbf{w}}_k\|. \tag{6}$$

From the learned cluster membership information, the corresponding $\mathbf{C}_{\mathbf{z}}$ matrix is associated with each noisy patch. This we denote as $\widehat{\mathbf{C}}_{\mathbf{z}_i}$. Hence, we estimate a covariance matrix for each patch, thereby avoiding the need to perform clustering or having to directly estimate $\mathbf{C}_{\mathbf{z}}$ from the noisy image patches. Later, in Sec. 4, we show that using such a method,

we can learn the patchwise $\widehat{\mathbf{C}}_{\mathbf{z}_i}$ matrix accurately enough to predict the denoising bound for any given noisy image.

3.2. Estimating the FIM

The other term that needs to be estimated to compute the bound is the FIM. Assuming that the corrupting noise is additive white Gaussian (AWG) with a known (or estimated) pixelwise standard deviation (σ), we need to estimate only the N_i values for each patch to estimate the FIM having the form of Eq. 3. In [1], this was estimated using the similarity measure of Eq. 4. However, this cannot be directly used for the noisy case. Nor is it possible to use the noise-free image database since we need to estimate the number of patches within the given *noisy* image that are similar to each noisy patch. A possible solution could be to use a denoised version of the noisy image to compute N_i. However, such an estimate would depend largely on the denoising algorithm. Moreover, the denoised image may well be over-smoothed and lack finer details making the estimation of N_i inaccurate. Thus we restrict ourselves to search for similar patches directly from the given noisy image. In such a case, the measure of similarity between noisy patches needs to be refined. This can be done by extending Eq. 4 as

$$\begin{aligned} & \mathbf{z}_j = \mathbf{z}_i + \boldsymbol{\varepsilon}_{ij} \\ \Rightarrow\ & \mathbf{y}_j - \boldsymbol{\eta}_j = \mathbf{y}_i - \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{ij} \\ \Rightarrow\ & \mathbf{y}_j = \mathbf{y}_i + \underbrace{\left(\boldsymbol{\eta}_j - \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{ij}\right)}_{\widetilde{\boldsymbol{\varepsilon}}_{ij}}, \end{aligned} \tag{7}$$

$$\text{where } \|\widetilde{\boldsymbol{\varepsilon}}_{ij}\|^2 = \|\boldsymbol{\varepsilon}_{ij}\|^2 + \|\boldsymbol{\eta}_j - \boldsymbol{\eta}_i\|^2 + 2\boldsymbol{\varepsilon}_{ij}^T(\boldsymbol{\eta}_j - \boldsymbol{\eta}_i)$$

$$\Rightarrow\ E[\|\widetilde{\boldsymbol{\varepsilon}}_{ij}\|^2] = E[\|\boldsymbol{\varepsilon}_{ij}\|^2] + 2\sigma^2 n, \tag{8}$$

since $\boldsymbol{\varepsilon}_{ij}$ is independent of the noise, and n is the number of pixels in each patch. Thus, any noisy patch $\mathbf{y}_j$ can be said to be similar to $\mathbf{y}_i$ if

$$\|\widetilde{\boldsymbol{\varepsilon}}_{ij}\|^2 \leqslant \gamma + \gamma_\eta, \tag{9}$$

where γ is defined in Eq. 4 and $\gamma_\eta = 2\sigma^2 n$ depends on the noise variance. Thus, by design, the new threshold for identifying similar patches across the image has a part γ that accounts for the difference in patches in the (unknown) original image and another γ_η for the difference arising due to noise corruption.

Once we have a mechanism of estimating N_i (denoted by $\hat{N}_i$) and $\widehat{\mathbf{C}}_{\mathbf{z}_i}$ for each patch in the noisy image, we compute the bound on the MSE for the entire image as

$$E[\|\mathbf{z}_i - \widehat{\mathbf{z}}_i\|^2] \geqslant \frac{1}{M}\sum_{i=1}^{M} \mathrm{Tr}\left[\left(\hat{N}_i\frac{\mathbf{I}}{\sigma^2} + \widehat{\mathbf{C}}_{\mathbf{z}_i}^{-1}\right)^{-1}\right]. \tag{10}$$

In the next section we provide experimental verification of our proposed method for estimating the bound.

(a) Some database images (b) Test images

Fig. 3. Few representative images from the database along with the test images that we will use to validate the proposed bounds estimation method.

4. EXPERIMENTAL RESULTS

In this section we show experimental results that illustrate how well the bound can be predicted from a noisy image for various levels of AWG noise. As explained in Sec. 3, to estimate $\widehat{\mathbf{C}}_{\mathbf{z}_i}$ we make use of a database of noise-free images. Naturally, the test images used in our experiments (Fig. 3(b)) are not included in the noise-free image database. In each case, the bounds estimated using the proposed approach is compared to the bounds computed from the noise-free image (ground truth), as outlined in [1].

As a first step of verification, we use a restrictive case where the images in the database belong to the same class as the noisy image. That is to say that if we are given a noisy image of a human face, we use only other face images to form our database. This has the advantage of making an accurate estimation of the bound while using patches from fewer images to form the database. We can also make use of fewer clusters (we use $K = 10$), resulting in a speedier estimation. However, this comes at the cost of human intervention in selecting some group of images similar in content to the given noisy image.

Hence it is more practical to use a generic database consisting of images from a vast variety of image classes. Fig. 3 shows a few representative images from which patches are used to make up our image database. The patches are preclustered into $K = 25$ clusters and $\mathbf{C}_\mathbf{z}$ is computed for each cluster. Note that another advantage of using a vast repository of patches is that it allows us to use a large K while also ensuring the presence of enough patches within each cluster to be able to compute $\mathbf{C}_\mathbf{z}$ stably.

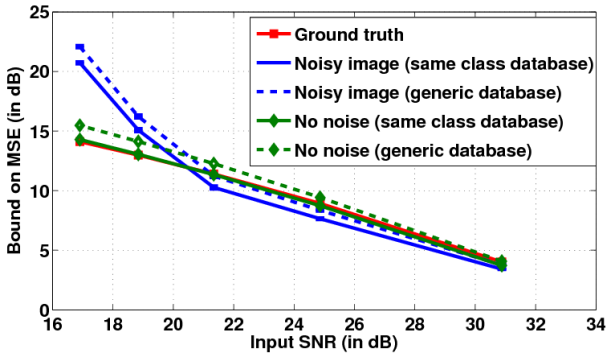
We first evaluate how well the covariance matrix is estimated for the noisy image patches. For this, we compute the bounds using the estimated $\widehat{\mathbf{C}}_{\mathbf{z}_i}$ for each patch in the noisy image. However, the N_i values are calculated from the noise-free image. Thus the differences in the predicted bounds are

purely due to errors in learning the covariance matrix. Fig. 4 shows experimental results for the face and mountain images for various signal-to-noise ratios (SNR) where the noise is AWG. There it can be seen that the bound estimated using the learned covariances are numerically similar to those estimated from the noise-free image, even for lower input SNRs. This is true irrespective of our choice of the noise-free database, that is, when the database is made up of fewer *same class* images as well as for the more generic database composed of a wider variety of images. This proves that using a database of noise-free images allows us to learn the covariances quite accurately.

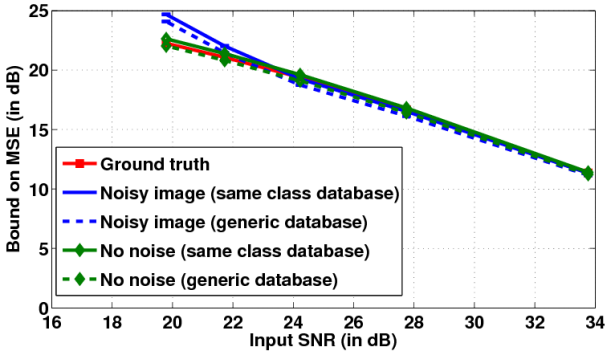
Next we consider estimating both $\widehat{\mathbf{C}}_{\mathbf{z}_i}$ and N_i as outlined in Sec. 3. Fig. 4(a) shows the bound estimated for various input SNR levels of the face image corrupted by AWG noise. There we can see that even for quite low input SNR (22 dB) the bound is estimated quite accurately in comparison to the bound computed from the noise-free face image. Good estimates for the bounds are obtained for both types of databases, although the generic database seems to perform better. A similar result can be seen in Fig. 4(b) where we compare the bounds estimated from the noisy mountain image to those estimated from the noise-free one. There too one can see that the bound estimation is quite accurate, even for a low SNR of 22dB (which corresponds to $\sigma \approx 15$). However, for stronger noise, the estimation is not consistent with the noise-free one due to difficulty in correctly estimating N_i values. This can be inferred from the *no noise* plots where the bound estimates when N_i values are estimated from noise-free images are very similar to the ones obtained entirely from the noise-free image. As a result, the bounds estimated for input SNR below 22dB are much larger than what is predicted from using the ground truth. Thus, although our bounds estimation scheme predicts an accurate bound from the noisy image for reasonable noise levels, it cannot be used if the noise is too strong.

5. CONCLUSIONS

In this paper we introduced a mechanism of estimating the bounds on the MSE that can be expected in denoising any given noisy image. We estimated the bound by treating it as two separate problems of estimating the parameters of the bounds formulation, namely the covariance matrix and the measure of redundancy for each patch. We showed, through experimental verification, that estimation of the bound is quite accurate when reasonable levels of noise are considered. Our results also illustrate that while the patchwise covariance matrix can be estimated robustly through the use of a collection of noise-free images, estimating patch redundancy (N_i) across the image is difficult for low SNR. This leads to inaccurate bound estimates when the corrupting noise is strong. Use of a partially denoised image to estimate the N_i values may possibly alleviate this problem. We consider this to be a direction for our continuing research.



(a) Face image



(b) Mountain image

Fig. 4. Bounds (in $10\log_{10}$ scale) estimated using a noise-free image database. The *noisy* bounds are those obtained using the proposed method; the *no noise* plots represent the case when N_i values are estimated from the noise-free image but $\widehat{\mathbf{C}}_{\mathbf{z}_i}$ is estimated from the look-up table; and *ground truth* denotes bounds estimated from the noise-free image [1].

6. REFERENCES

[1] P. Chatterjee and P. Milanfar, "Is Denoising Dead?," *IEEE Trans. on Image Proc.*, vol. 19, no. 4, April 2010.

[2] T. Treibitz and Y. Y. Schechner, "Recovery Limits in Pointwise Degradation," in *Proc. of IEEE Intl. Conf. on Computational Photography*, San Francisco, USA, April 2009.

[3] S. Voloshynovskiy, O. Koval, and T. Pun, "Image Denoising Based on the Edge-Process Model," *Signal Processing*, vol. 85, no. 10, pp. 1950–1969, October 2005.

[4] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Trans. on Image Proc.*, vol. 16, no. 2, pp. 349–66, February 2007.

[5] P. Chatterjee and P. Milanfar, "Clustering-Based Denoising with Locally Learned Dictionaries," *IEEE Trans. on Image Proc.*, vol. 18, no. 7, pp. 1438–1451, July 2009.

[6] B. Efron, "Bootstrap Methods: Another Look at the Jacknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[7] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.