IMAGE DENOISING BY TARGETED EXTERNAL DATABASES

Enming Luo[1], Stanley H. Chan [2], and Truong Q. Nguyen[1]

[1]University of California, San Diego, Dept of ECE, 9500 Gilman Drive, La Jolla, CA 92093.
[2]Harvard School of Engineering and Applied Sciences, 33 Oxford Street, Cambridge, MA 02138.

ABSTRACT

Classical image denoising algorithms based on single noisy images or generic image databases will soon reach their performance limits. In this paper, we propose to consider denoising using *targeted* external image databases. Formulating denoising as an optimal filter design problem, we utilize the targeted databases to (1) determine the basis functions of the optimal filter by means of group sparsity; (2) determine the spectral coefficients of the optimal filter by means of localized priors. For a variety of scenarios such as text images, multiview images, and face images, we demonstrate superior denoising results over existing algorithms.

***Index Terms*—** Patch-based denoising, group sparsity, Bayesian minimum mean squared error, external database, optimal filter

1. INTRODUCTION

Patch-based image denoising algorithm [1, 2, 3, 4, 5] refers to a class of recently developed denoising methods based on the concept of patch similarity. For an $\sqrt{n} \times \sqrt{n}$ patch $\boldsymbol{q} \in \mathbb{R}^n$ of the noisy image, a patch-based algorithm finds a set of similar patches $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k \in \mathbb{R}^n$, and applies some linear (or non-linear) function φ to obtain an estimated (denoised) patch $\widehat{\boldsymbol{p}_0}$ as

$$\widehat{\boldsymbol{p}_0} = \varphi(\boldsymbol{q};\, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_k). \tag{1}$$

For example, the non-local means [1] considers φ as a weighted averaging function, whereas the BM3D [2] considers φ as a transform-shrinkage operation.

In applying the patch-based denoising algorithm, finding the similar patches $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$ is the key. Typically, there are two sources of these patches: the noisy image itself and an external database. Finding similar patches from the noisy image itself is more popular because many patches tend to recur within the image [6, 7] However, single image denoising is also known to have limited performance, especially for rare patches [8]. This motivates the use of external databases [9, 10, 11]. In [12], it is shown that performance limit of image denoising is achievable only by using external databases. However, most of the existing external denoising algorithms use *generic* databases, in the sense that no prior knowledge about the scene is used. This raises a natural question: are there situations under which *targeted* databases can be utilized to improve the denoising quality?

In fact, building targeted databases is plausible in many scenarios. As will be illustrated in later parts of this paper, targeted databases can easily be built for text images (e.g., newspapers and documents), human faces (under certain conditions), and images captured by multi-view camera systems. Other possible scenarios include: images of licence plates, medical CT and MRI images, and images of landmarks.

Assuming that these databases are given, one fundamental question to ask is: what is the corresponding denoising algorithm? In other words, is it possible to design a computationally efficient denoising procedure that can *maximally* utilize the databases? The goal of this paper is to provide an answer to this question by showing that for the above mentioned applications, an algorithm can be designed to outperform many existing methods by a significant margin.

1.1. Related Work

The focus of this paper is about denoising algorithms using external databases. In general, there are two directions in the literature that are relevant to our work. The first approach is to modify existing algorithms to handle external databases by brute-force extensions. For example, one can always feed databases to a multiframe algorithm, *e.g.* [13, 14, 15, 16], to extend the methods for this case. However, such approach is rather ad-hoc and there is no theoretical guarantee on the performance.

The other approach is to learn the prior of the database and denoise the image using a maximum a posteriori (MAP) estimation *e.g.* [4, 17, 18, 9, 19]. Some of these methods have performance guarantee, yet a large number of samples are needed for training the priors. In practice, this is not always possible because finding a sufficient number of targeted images could be difficult.

1.2. Contributions

In contrast to the existing methods, the proposed algorithm requires only a few targeted images in the database. Moreover, the proposed algorithm offers two new insights into the denoising problem.

First, we show that when designing a linear denoising filter, the basis matrix can be learned by minimizing a convex problem involving group sparsity, and the solution is the classical eigen-decomposition. This provides justifications of many well-known denoising algorithms in which PCA is used as a learning step. Second, we show that when estimating the spectral components of the denoising filter, it is possible to improve the robustness by introducing local priors. By minimizing the associated Bayesian mean squared error, the denoising quality can be improved.

The rest of the paper is organized as follows. In Section 2 we present the problem setup and the proposed algorithm. Experimental results are shown in Section 3, and a concluding remark is given in Section 4. Technical details of this paper will be presented in a follow-up journal paper.

Contact author: E. Luo, e-mail: eluo@ucsd.edu. The work of the first and last authors is supported, in part, by a NSF grant CCF-1065305. The work of the second author is supported, in part, by a Croucher Foundation post-doctoral research fellowship.

2. PROPOSED METHOD

In this section, we first provide a brief review of the classical optimal denoising filter design problem and highlight its limitations. Consequently, we describe the proposed method and discuss its relation to existing methods.

2.1. Optimal Denoising Filter

We consider the denoising task as an optimal filter design problem for its simplicity and analytic tractability [20]. Given a clean patch $\boldsymbol{p}_0 \in \mathbb{R}^n$, we model the observed noisy patch as $\boldsymbol{q} = \boldsymbol{p}_0 + \boldsymbol{n}$, where $\boldsymbol{n}$ is a vector of i.i.d. Gaussian noise of zero mean and variance σ^2. The optimal denoising filter problem is to find a linear operator $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ such that an estimate $\widehat{\boldsymbol{p}_0}$ can be obtained by $\widehat{\boldsymbol{p}_0} = \boldsymbol{A}\boldsymbol{q}$. Here, we assume that $\boldsymbol{A}$ is symmetric, or otherwise the Sinkhorn-Knopp iteration [21] can be used to symmetrize $\boldsymbol{A}$.

Since $\boldsymbol{A}$ is symmetric, it is valid to apply the eigen-decomposition: $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, for some eigenvectors $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n] \in \mathbb{R}^{n\times n}$ and eigenvalues $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_n\} \in \mathbb{R}^{n\times n}$. Therefore, the filter design problem becomes the question of finding $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ so that the linear estimate $\widehat{\boldsymbol{p}_0} = \boldsymbol{A}\boldsymbol{q}$ has the minimum MSE compared to $\boldsymbol{p}_0$:

$$(\boldsymbol{U}, \boldsymbol{\Lambda}) = \underset{\boldsymbol{U},\boldsymbol{\Lambda}}{\text{argmin}} \; \mathbb{E}\left[\left\|\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\boldsymbol{q} - \boldsymbol{p}_0\right\|_2^2\right], \tag{2}$$

subject to the constraint that $\boldsymbol{U}$ is an orthonormal matrix. The solution of this problem is given as follows.

Lemma 1. *The estimate $\widehat{\boldsymbol{p}_0}$ under the optimal solution of* (2) *is*

$$\widehat{\boldsymbol{p}_0} = \boldsymbol{U}\left(\text{diag}\left\{\frac{\|\boldsymbol{p}_0\|^2}{\|\boldsymbol{p}_0\|^2 + \sigma^2}, 0, \ldots, 0\right\}\right)\boldsymbol{U}^T\boldsymbol{q},$$

where $\boldsymbol{U}$ is any orthonormal matrix with $\boldsymbol{u}_1 = \boldsymbol{p}_0/\|\boldsymbol{p}_0\|$.

Evidently, this (trivial) solution is never achievable because it involves the ground truth $\boldsymbol{p}_0$. Therefore, pre-defined basis $\boldsymbol{U}$ are commonly used in replacing (2), *e.g.* Fourier basis [2] or PCA basis [3, 5]. However, the optimality of these basis are not fully understood, especially in the external database situation. Additionally, even when $\boldsymbol{U}$ is determined, the spectrum $\boldsymbol{\Lambda}$ is still not known. This motivates us to consider the following design processes.

2.2. Optimal $\boldsymbol{U}$ by Group Sparsity

Determining $\boldsymbol{U}$ is an ill-posed problem because any $\boldsymbol{U}$ satisfying the orthogonality condition $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$ is a valid solution. Our proposed method constructs $\boldsymbol{U}$ by exploiting structures of the targeted database. First, we note that for a given targeted database, the best k matching patches $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$ must be highly similar. Consequently, if we consider the matrix $\boldsymbol{P} \overset{\text{def}}{=} [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k]$, then a good basis $\boldsymbol{U}$ must be the one that preserves the similarity in the spectral domain. To this end, we consider the following group sparsity optimization problem:

$$\begin{aligned} \underset{\boldsymbol{U}}{\text{minimize}} \quad & \|\boldsymbol{U}^T\boldsymbol{P}\|_{1,2} \\ \text{subject to} \quad & \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}. \end{aligned} \tag{3}$$

The $\ell_{1,2}$-norm in (3) is a measure of group sparsity that enforces $\boldsymbol{U}^T\boldsymbol{P}$ to have similar magnitudes locating at similar positions. This notion of group sparsity has been previously used in [4], but towards a different end. A more interesting yet less known fact about (3) is that the solution is the simple eigen-decomposition:

Lemma 2. *The solution to* (3) *is*

$$[\boldsymbol{U}, \boldsymbol{S}] = eig(\boldsymbol{P}\boldsymbol{P}^T), \tag{4}$$

where $\boldsymbol{U}$ is the eigenvector matrix, and $\boldsymbol{S}$ is the eigenvalue matrix.

The result of Lemma 2 implies that many existing PCA-based methods (e.g., [5][14]) indeed assume a group sparsity prior, although implicitly. If we further define diagonal weight matrices $\boldsymbol{W}_1 \in \mathbb{R}^{n\times n}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{k\times k}$, and consider the problem

$$\begin{aligned} \underset{\boldsymbol{U}}{\text{minimize}} \quad & \|\boldsymbol{U}^T\boldsymbol{W}_1^{1/2}\boldsymbol{P}\boldsymbol{W}_2^{1/2}\|_{1,2} \\ \text{subject to} \quad & \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}, \end{aligned} \tag{5}$$

then the solution becomes a generalization of the shape-adaptive BM3D-PCA [3] where the spatial shape-adaptivity is controlled by $\boldsymbol{W}_1$ and the patch intensity similarity is controlled by $\boldsymbol{W}_2$. In the rest of this paper, we let $\boldsymbol{W}_1 = \boldsymbol{I}$, and define

$$\boldsymbol{W}_2 = \frac{1}{\eta}\,\text{diag}\left\{e^{-\|\boldsymbol{q}-\boldsymbol{p}_1\|^2/h^2}, \ldots, e^{-\|\boldsymbol{q}-\boldsymbol{p}_k\|^2/h^2}\right\}, \tag{6}$$

for some decay parameter h and normalization constant η. Such choice of $\boldsymbol{W}_2$ is to ensure that dissimilar patches have less contribution in computing $\boldsymbol{U}$.

2.3. Optimal $\boldsymbol{\Lambda}$ by Localized Prior

Given $\boldsymbol{U}$, one naive approach to determine $\boldsymbol{\Lambda}$ is to solve (2) for fixed $\boldsymbol{U}$. In this case, it is not difficult to show that the optimal solution is

$$\widehat{\boldsymbol{p}_0} = \boldsymbol{U}\left(\text{diag}\left\{\frac{(\boldsymbol{u}_1^T\boldsymbol{p}_0)^2}{(\boldsymbol{u}_1^T\boldsymbol{p}_0)^2 + \sigma^2}, \ldots, \frac{(\boldsymbol{u}_n^T\boldsymbol{p}_0)^2}{(\boldsymbol{u}_n^T\boldsymbol{p}_0)^2 + \sigma^2}\right\}\right)\boldsymbol{U}^T\boldsymbol{q}. \tag{7}$$

However, since we never have access to the ground truth $\boldsymbol{p}_0$, the spectral components $\lambda_i = \frac{(\boldsymbol{u}_i^T\boldsymbol{p}_0)^2}{(\boldsymbol{u}_i^T\boldsymbol{p}_0)^2+\sigma^2}$ ($i = 1, \ldots, n$) have to be approximated, e.g., using an initial estimate via transform-shrinkage [2, 3]: $\lambda_i = \frac{(\boldsymbol{u}_i^T\overline{\boldsymbol{p}_0})^2}{(\boldsymbol{u}_i^T\overline{\boldsymbol{p}_0})^2+\sigma^2}$ for some estimate $\overline{\boldsymbol{p}_0}$, or using the noisy estimate [5]: $\lambda_i = \frac{(\boldsymbol{u}_i^T\boldsymbol{q})^2-\sigma^2}{(\boldsymbol{u}_i^T\boldsymbol{q})^2}$. Yet, regardless of any particular choice of methods, the estimate is a *point* estimate. As a result, the denoising quality is sensitively depending on which point is chosen. Uncertainty of the estimate is never taken into account.

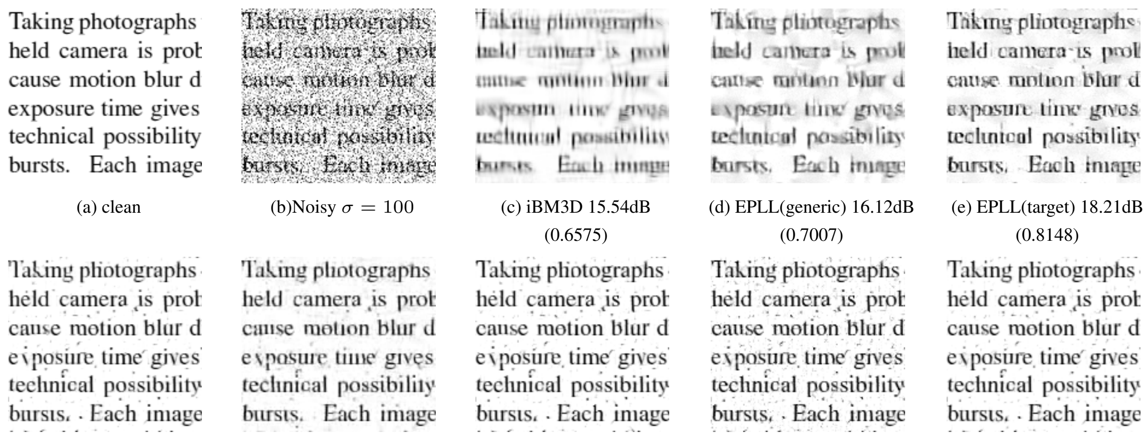
Our proposed method includes uncertainty when estimating $\boldsymbol{\Lambda}$. The idea is to assume a prior distribution and consider the Bayesian mean squared error (BMSE):

$$\text{BMSE} = \mathbb{E}_{\boldsymbol{p}}\left[\mathbb{E}_{\boldsymbol{q}|\boldsymbol{p}}\left[\left\|\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\boldsymbol{q} - \boldsymbol{p}\right\|_2^2 \;\middle|\; \boldsymbol{p}\right]\right]. \tag{8}$$

In evaluating (8), it is important to specify the conditional distribution $f(\boldsymbol{q} \mid \boldsymbol{p})$ and the prior $f(\boldsymbol{p})$.

The conditional distribution $f(\boldsymbol{q} \mid \boldsymbol{p})$ is the i.i.d. Gaussian due to the definition of the noise model. Thus, $f(\boldsymbol{q} \mid \boldsymbol{p}) = \mathcal{N}(\boldsymbol{p}, \sigma^2\boldsymbol{I})$. The prior distribution $f(\boldsymbol{p})$ is defined based on the trustfulness of the k matching patches $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$. Letting $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the first and second cumulant of $f(\boldsymbol{p})$, we can show that the minimum BMSE is achieved at

$$\boldsymbol{\Lambda} = \underset{\boldsymbol{\Lambda}}{\text{argmin}}\;\text{BMSE} = \frac{\text{diag}\left\{\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U} + \boldsymbol{U}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{U}\right\}}{\text{diag}\left\{\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U} + \boldsymbol{U}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{U}\right\} + \sigma^2\boldsymbol{I}}, \tag{9}$$



(a) clean (b)Noisy $\sigma = 100$ (c) iBM3D 15.54dB (0.6575) (d) EPLL(generic) 16.12dB (0.7007) (e) EPLL(target) 18.21dB (0.8148)

(a) eNLM 20.81dB (0.8636) (b) eBM3D 20.07dB (0.8376) (c) eBM3D-PCA 21.32dB (0.8611) (d) eLPG-PCA 20.25dB (0.7670) (d) ours **21.87dB** (**0.8639**)

Fig. 1. Text image denoising: visual comparison and objective comparison (PSNR and SSIM in the parenthesis)

where the division is element-wise. Note that in computing (9), we do not need a specific form for $f(\boldsymbol{p})$ except the knowledge of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

To define $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we first note that $\boldsymbol{\mu}$ is the desired mean of the prior and $\boldsymbol{\Sigma}$ is the covariance measuring the uncertainty. Based on the set of k similar patches $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$, we define

$$\boldsymbol{\mu} = \sum_{i=1}^{k} \omega_i \boldsymbol{p}_i, \quad \boldsymbol{\Sigma} = \sum_{i=1}^{k} \omega_i (\boldsymbol{p}_i - \boldsymbol{\mu})(\boldsymbol{p}_i - \boldsymbol{\mu})^T, \qquad (10)$$

where $\omega_i = \frac{1}{\eta} e^{-\|\boldsymbol{q}-\boldsymbol{p}_i\|^2/h^2}$ is the weight defined in (6). Intuitively, $\boldsymbol{\mu}$ is the best-k non-local mean solution of the patches, and $\boldsymbol{\Sigma}$ is the covariance of the patches.

Our choice of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is computationally very efficient because of the following lemma.

Lemma 3. *Using $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ defined in* (10)*, the optimal $\boldsymbol{\Lambda}$ defined in* (9) *is given by*

$$\boldsymbol{\Lambda} = \frac{\boldsymbol{S}}{\boldsymbol{S} + \sigma^2 \boldsymbol{I}}, \qquad (11)$$

where $[\boldsymbol{U}, \boldsymbol{S}] = eig(\boldsymbol{P}\boldsymbol{W}_2\boldsymbol{P}^T)$ is the eigen-decomposition of the weighted matrix $\boldsymbol{P}\boldsymbol{W}_2^{1/2}$.

Therefore, by solving the group sparsity problem (5), we automatically obtain both $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$, simultaneously. The overall algorithm is given in Algorithm 1.

Algorithm 1 Proposed Algorithm

Input: noisy patch $\boldsymbol{q}$, and similar patches $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$
Output: estimate $\widehat{\boldsymbol{p}_0}$
Learn $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$

- Form data matrix $\boldsymbol{P}$ and weight matrix $\boldsymbol{W}_2$
- Compute eigen-decomposition $[\boldsymbol{U}, \boldsymbol{S}] = \mathrm{eig}(\boldsymbol{P}\boldsymbol{W}_2\boldsymbol{P}^T)$
- Compute $\boldsymbol{\Lambda} = \frac{\boldsymbol{S}}{\boldsymbol{S}+\sigma^2\boldsymbol{I}}$ (element-wise division)

Denoise: $\widehat{\boldsymbol{p}_0} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\boldsymbol{q}$

It is interesting to compare the proposed Bayesian approach with existing methods. First, we note that many PCA-based patch denoising algorithms [2, 3, 5, 14] can be generalized under our Bayesian framework. In those situations, the prior $f(\boldsymbol{p})$ is simply a delta function centered at the initial guess $\overline{\boldsymbol{p}_0}$: $f(\boldsymbol{p}) = \delta(\boldsymbol{p} - \overline{\boldsymbol{p}_0})$. As a comparison to the generic learning methods, such as [4, 9, 17, 22], our method is a *local* prior whereas theirs can be considered as a *global* prior. Learning a global prior requires a large number of samples in the training set, which could be difficult because getting *many* targeted images is not always possible. Even if the training sets are sufficiently large, the computing load is still intensive. In contrast, the proposed method assumes different priors $f(\boldsymbol{p})$ for different patches $\boldsymbol{p}_0$. As a result, our method is more advantageous because fewer samples are needed to train the prior.

3. EXPERIMENTAL RESULTS

3.1. Experiment Settings

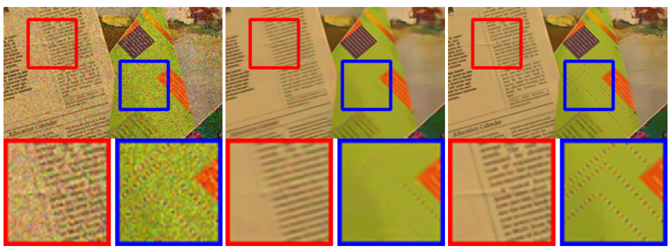
In this section we evaluate the performance of the proposed method by comparing to existing algorithms. The methods we considered in the comparison include BM3D[2], BM3D-PCA[3], LPG-PCA[5], NLM[1] and EPLL[4]. Except for EPLL, all other four methods are re-implemented to enlarge the patch search over several images. As for NLM, instead of using all patches in the database, we consider only the best k patches following [23]. For EPLL, we considered both the default patch prior learned from a generic database, and a new prior learned from our targeted database by running the same EM algorithm. To emphasize the difference between the original algorithms of the existing methods and the corresponding new implementations for external databases, we denote "i" (*internal*) for the single image denoising algorithms, and "e" (*external*) for the corresponding extension for external databases.

3.2. Denoising Text and Documents

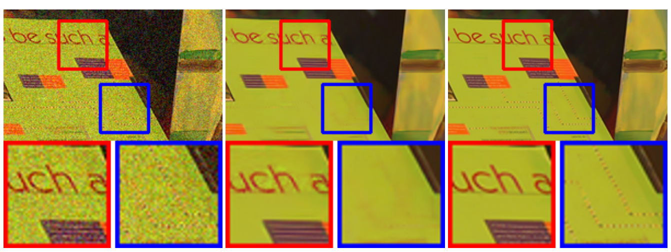
Our first experiment considers denoising a noisy text image using a collection of arbitrarily chosen text images of the same scale. The motivation of this experiment is to simulate denoising a document with the assistance of other similar but non-identical texts. The results are shown in Figure 1, where we observe that the proposed method yields the highest PSNR and SSIM values. It is evident that when the targeted database contains only a few samples (9 other text images in our experiment), existing training-based methods, such as EPLL, do not perform well.

3.3. Denoising Multiview Images

The second experiment considers the scenario of capturing images using a multiview camera system. Suppose that one or more cameras are not functioning properly so that some images are corrupted by noise. We demonstrate that with the aid of the remaining clean images, it is possible to recover the noisy images to a great extent. In Figure 2 we illustrate two simulation results of the experiment. For both simulations, one of the 5 multiview images is corrupted with i.i.d. Gaussian noise. Using the proposed method, we see that the noisy images are improved significantly. A PSNR comparison with existing methods is shown in Figure 3.



(a) noisy view (std:20) (b) iBM3D: 29.10dB (c) ours: 34.62dB



(d) noisy view (std:20) (e) iBM3D: 30.42dB (f) ours: 35.13dB

Fig. 2. Multiview image denoising using the proposed method and internal BM3D. [Top] "Barn"; [Bottom] "Venus".

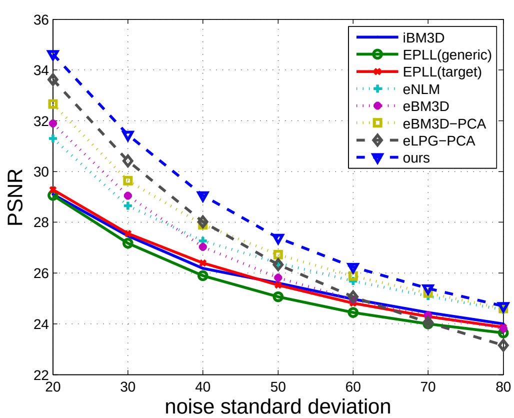


Fig. 3. Multiview image denoising for "Barn": PSNR vs noise levels. In this plot, each PSNR value is averaged over 8 independent trials to reduce the bias due to a particular noise pattern.

3.4. Denoising Human Faces

The third experiment considers denoising face images. In this experiment, we consider a set of face images from [24]. We simulate the denoising task by adding noise to a randomly chosen image and use other images in the database for denoising. Note that the faces are not pre-processed, in the sense that they have different expressions and alignments. However, even in the presence of this degree of image variations, the proposed method still performs satisfactorily over a range of noise levels. The results are shown in Figure 4 and Figure 5.

Fig. 4. Face denoising of Gore dataset [24]. [Top] Database images; [Bottom] Denoising results.

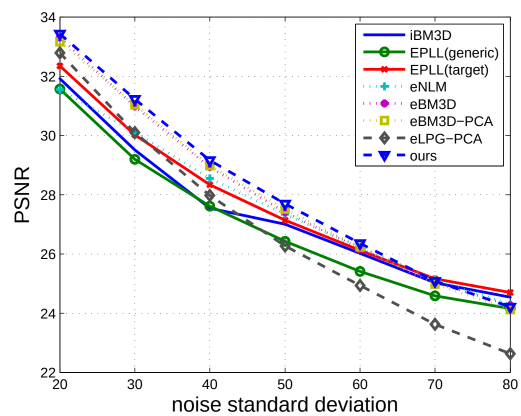


Fig. 5. Face denoising results: PSNR vs noise level. In this plot, each PSNR value is averaged over 8 independent trials to reduce the bias due to a particular noise pattern.

4. CONCLUSION

Classical image denoising methods based on single noisy input or generic databases are approaching their performance limits. We envision that future image denoising should be target-oriented, i.e., for specific objects to be denoised, the corresponding images should be used for training. To address this new paradigm shift in image denoising, we present algorithms and corresponding simulations of using targeted databases for optimal linear denoising filter design. Our proposed method, based on group sparsity and localized priors, showed robustness and performance superiority over a wide range of existing algorithms. Future work includes detailed sensitivity analysis of the algorithm.

5. REFERENCES

[1] A. Buades, B. Coll, and J. Morel, “A review of image denoising algorithms, with a new one,” *SIAM Multi. Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “BM3D image denoising with shape-adaptive principal component analysis,” in *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS ’09)*, 2009.

[4] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Non-local sparse models for image restoration,” in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR ’09)*, 2009, pp. 2272–2279.

[5] 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, pp. 1531–1549, Apr. 2010.

[6] M. Zontak and M. Irani, “Internal statistics of a single natural image,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR’11)*, 2011, pp. 977–984.

[7] I. Mosseri, M. Zontak, and M. Irani, “Combining the power of internal and external denoising,” in *Proc. Intl. Conf. Computational Photography (ICPR ’13)*, 2013.

[8] P. Chatterjee and P. Milanfar, “Is denoising dead?,” *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 895–911, Apr. 2010.

[9] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[10] N. Joshi, W. Matusik, E.H. Adelson, and D.J. Kriegman, “Personal photo enhancement using example images,” *ACM Trans. Graphics*, vol. 29, no. 2, pp. 1–15, 2010.

[11] S.H. Chan, T. Zickler, and Y.M. Lu, “Fast non-local filtering by random sampling: it works, especially for large images,” in *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Process. (ICASSP ’13)*, 2013.

[12] A. Levin and B. Nadler, “Natural image denoising: Optimality and inherent bounds,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR’11)*, 2011, pp. 2833–2840.

[13] K. Dabov, A. Foi, and K. Egiazarian, “Video denoising by sparse 3D transform-domain collaborative filtering,” in *Proc. European Signal Processing Conference*, 2007, pp. 145–149.

[14] L. Zhang, S. Vaddadi, H. Jin, and S. Nayar, “Multiple view image denoising,” in *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR’09)*, 2009, pp. 1542–1549.

[15] E. Luo, S.H. Chan, S. Pan, and T.Q. Nguyen, “Adaptive non-local means for multiview image denoising: Searching for the right patches via a statistical approach,” in *Proc. IEEE Intl. Conf. Image Process. (ICIP’13)*, 2013.

[16] W. Dong, L. Zhang, G. Shi, and X. Li, “Nonlocally centralized sparse representation for image restoration,” *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[17] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in *Proc. IEEE Intl. Conf. Computer Vision (ICCV’11)*, Nov. 2011, pp. 479–486.

[18] S. Roth and M. Black, “Fields of experts: A framework for learning image priors,” in *in Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition(CVPR’05), 2005*. IEEE, 2005, vol. 2, pp. 860–867.

[19] L. Sun and J. Hays, “Super-resolution from internet-scale scene matching,” in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, 2012.

[20] P. Milanfar, “A tour of modern image filtering,” *IEEE Signal Processing Magazine*, vol. 30, pp. 106–128, Jan. 2013.

[21] P. Milanfar, “Symmetrizing smoothing filters,” *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 263–284, 2013.

[22] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: Design of dictionaries for sparse representation,” in *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS ’05)*, 2005, vol. 5, pp. 9–12.

[23] J. Kervrann and J. Boulanger, “Local adaptivity to variable smoothness for exemplar-based image regularization and representation,” *Intl. J. Computer Vision*, vol. 79, no. 1, pp. 45–69, 2008.

[24] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images,” in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR’10)*, 2010, pp. 763–770.