

IMAGE DENOISING BY TARGETED EXTERNAL DATABASES

Enming Luo¹, Stanley H. Chan², and Truong Q. Nguyen¹

¹University of California, San Diego, Dept of ECE, 9500 Gilman Drive, La Jolla, CA 92093.

²Harvard School of Engineering and Applied Sciences, 33 Oxford Street, Cambridge, MA 02138.

ABSTRACT

Classical image denoising algorithms based on single noisy images and generic image databases will soon reach their performance limits. In this paper, we propose to denoise images 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 algorithms [1, 2, 3, 4, 5] refer to a class of recently developed denoising methods based on the concept of patch similarity. For a $\sqrt{n} \times \sqrt{n}$ patch $\mathbf{q} \in \mathbb{R}^n$ of the noisy image, a patch-based algorithm finds a set of similar patches $\mathbf{p}_1, \dots, \mathbf{p}_k \in \mathbb{R}^n$, and applies some linear (or non-linear) function φ to obtain an estimated (denoised) patch $\hat{\mathbf{p}}$ as

$$\hat{\mathbf{p}} = \varphi(\mathbf{q}; \mathbf{p}_1, \dots, \mathbf{p}_k). \quad (1)$$

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

In applying patch-based denoising algorithms, finding similar patches $\mathbf{p}_1, \dots, \mathbf{p}_k$ 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 patches tend to recur within the image [6, 7]. However, this approach has limited performance, especially for rare patches [8]. Another source of obtaining similar patches is to use external databases [9, 10, 11, 12], which in theory can achieve the minimum mean squared estimation error [13]. 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 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 the targeted databases are given, one fundamental question is: what is the corresponding denoising algorithm? Or 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 and its performance is better than several existing methods.

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 modify existing single image denoising algorithms, *e.g.*, [1, 2, 3, 14], so that they search similar patches from an external database. Similarly, one can treat a database as “videos” for multi-image denoising algorithms, *e.g.*, [15, 16, 17, 18]. However, both approaches are heuristic in which 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 method, *e.g.*, [4, 9, 19, 20, 21]. While some of these methods have performance guarantee, a large number of samples are needed for training the priors which are not always available in practice.

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 solving a convex optimization 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, a localized prior can be used and the denoising quality is improved by minimizing the associated Bayesian mean squared error.

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 E. Luo and T. Q. Nguyen is supported, in part, by a NSF grant CCF-1065305. The work of S. H. Chan is supported, in part, by a Croucher Foundation post-doctoral research fellowship.

2. PROPOSED METHOD

We first provide a brief review of the classical optimal denoising filter design problem and highlight its limitations. Then, 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 [22, 23, 24]. Given a clean patch $\mathbf{p} \in \mathbb{R}^n$, we model the observed noisy patch as $\mathbf{q} = \mathbf{p} + \mathbf{n}$, where \mathbf{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 $\mathbf{A} \in \mathbb{R}^{n \times n}$ such that an estimate $\hat{\mathbf{p}}$ can be obtained by $\hat{\mathbf{p}} = \mathbf{A}\mathbf{q}$. Here, we assume that \mathbf{A} is symmetric, or otherwise the Sinkhorn-Knopp iteration [25] can be used to symmetrize \mathbf{A} .

Since \mathbf{A} is symmetric, it is valid to apply the eigen-decomposition, $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, to obtain the eigenvectors $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ and the eigenvalues $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_n\} \in \mathbb{R}^{n \times n}$. Therefore, the filter design problem becomes the question of finding \mathbf{U} and $\mathbf{\Lambda}$ so that the linear estimate $\hat{\mathbf{p}} = \mathbf{A}\mathbf{q}$ has the minimum mean squared error (MSE) compared to \mathbf{p} :

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

subject to the constraint that \mathbf{U} is an orthonormal matrix. The optimal solution of (2) yields the following denoising result:

Lemma 1. *The denoised patch $\hat{\mathbf{p}}$ using the optimal \mathbf{U} and $\mathbf{\Lambda}$ of (2) is*

$$\hat{\mathbf{p}} = \mathbf{U} \left(\text{diag} \left\{ \frac{\|\mathbf{p}\|^2}{\|\mathbf{p}\|^2 + \sigma^2}, 0, \dots, 0 \right\} \right) \mathbf{U}^T \mathbf{q},$$

where \mathbf{U} is any orthonormal matrix with the first column $\mathbf{u}_1 = \mathbf{p}/\|\mathbf{p}\|$.

Evidently, this (trivial) solution is not achievable because it involves the ground truth \mathbf{p} . Therefore, many existing methods use pre-defined bases, e.g. Fourier basis [2] and PCA basis [3, 5], to fix \mathbf{U} in (2). However, the optimality of these bases are not fully understood. Moreover, even if \mathbf{U} is pre-defined, an optimal choice of the spectrum $\mathbf{\Lambda}$ which does not depend on \mathbf{p} is yet to be defined. This motivates the following design procedure.

2.2. Optimal \mathbf{U} by Group Sparsity

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

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

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

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

$$[\mathbf{U}, \mathbf{S}] = \text{eig}(\mathbf{P}\mathbf{P}^T), \quad (4)$$

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

The result of Lemma 2 implies that many existing PCA-based methods (e.g., [5, 16]) indeed assume a group sparsity prior, although implicitly. If we further define diagonal weight matrices $\mathbf{W}_1 \in \mathbb{R}^{n \times n}$ and $\mathbf{W}_2 \in \mathbb{R}^{k \times k}$, and consider the problem

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

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

$$\mathbf{W}_2 = \frac{1}{\eta} \text{diag} \left\{ e^{-\|\mathbf{q}-\mathbf{p}_1\|^2/h^2}, \dots, e^{-\|\mathbf{q}-\mathbf{p}_k\|^2/h^2} \right\}, \quad (6)$$

for some decay parameter h and normalization constant η . Such choice of \mathbf{W}_2 ensures that dissimilar patches have less contribution in computing \mathbf{U} .

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

For fixed \mathbf{U} , it is not difficult to show that the optimal denoised patch is

$$\hat{\mathbf{p}} = \mathbf{U} \left(\text{diag} \left\{ \frac{(\mathbf{u}_1^T \mathbf{p})^2}{(\mathbf{u}_1^T \mathbf{p})^2 + \sigma^2}, \dots, \frac{(\mathbf{u}_n^T \mathbf{p})^2}{(\mathbf{u}_n^T \mathbf{p})^2 + \sigma^2} \right\} \right) \mathbf{U}^T \mathbf{q}. \quad (7)$$

However, since we do not have access to the ground truth \mathbf{p} , the spectral components $\lambda_i \stackrel{\text{def}}{=} \frac{(\mathbf{u}_i^T \mathbf{p})^2}{(\mathbf{u}_i^T \mathbf{p})^2 + \sigma^2}$ ($i = 1, \dots, n$) have to be approximated. For example, in [2, 3], λ_i is approximated as $\lambda_i = \frac{(\mathbf{u}_i^T \bar{\mathbf{p}})^2}{(\mathbf{u}_i^T \bar{\mathbf{p}})^2 + \sigma^2}$, where $\bar{\mathbf{p}}$ is an initial estimate using a denoising algorithm, e.g., BM3D. However, this type of approximations are suboptimal because λ_i is sensitive to the choice of $\bar{\mathbf{p}}$.

Our proposed method takes into account of the uncertainty of the initial estimate $\bar{\mathbf{p}}$ when estimating $\mathbf{\Lambda}$. The idea is to assume a prior distribution of the patch \mathbf{p} and minimize the Bayesian mean squared error (BMSE) over $\mathbf{\Lambda}$:

$$\text{BMSE} = \mathbb{E}_{\mathbf{p}} \left[\mathbb{E}_{\mathbf{q}|\mathbf{p}} \left[\left\| \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \mathbf{q} - \mathbf{p} \right\|_2^2 \mid \mathbf{p} \right] \right]. \quad (8)$$

In (8), the conditional distribution $f(\mathbf{q} \mid \mathbf{p})$ is i.i.d. Gaussian, following from the definition of the noise model. Thus, $f(\mathbf{q} \mid \mathbf{p}) = \mathcal{N}(\mathbf{p}, \sigma^2 \mathbf{I})$. The prior distribution $f(\mathbf{p})$ is defined according to the trustfulness of the k matching patches $\mathbf{p}_1, \dots, \mathbf{p}_k$. Here, we assume that the exact expression of $f(\mathbf{p})$ is unknown, but the mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ of $f(\mathbf{p})$ can be reasonably estimated. Consequently, for a given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the minimum BMSE is achieved at

$$\mathbf{\Lambda} = \underset{\mathbf{\Lambda}}{\text{argmin}} \text{BMSE} = \frac{\text{diag} \{ \mathbf{U}^T \boldsymbol{\Sigma} \mathbf{U} + \mathbf{U}^T \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{U} \}}{\text{diag} \{ \mathbf{U}^T \boldsymbol{\Sigma} \mathbf{U} + \mathbf{U}^T \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{U} \} + \sigma^2 \mathbf{I}}, \quad (9)$$

where the division is element-wise.

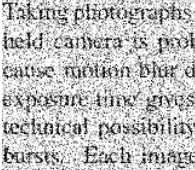
Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image		Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image	Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image	Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image
(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)
Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image	Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image	Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image	Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image	Taking photographs held camera is prot cause motion blur d exposure time gives technical possibility bursts. Each image
(f) eNLM 20.81dB (0.8636)	(g) eBM3D 20.07dB (0.8376)	(h) eBM3D-PCA 21.32dB (0.8611)	(i) eLPG-PCA 20.25dB (0.7670)	(j) ours 21.87dB (0.8639)

Fig. 1: Denoising text images: Visual comparison and objective comparison (PSNR and SSIM in the parenthesis). Prefix “i” stands for internal denoising (*i.e.*, single image denoising), and prefix “e” stands for external denoising (*i.e.*, using external databases).

It remains to specify $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Our choice of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is based on our belief of the true mean of the prior and the associated uncertainty of the belief. To this end, we define

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

where $\omega_i = \frac{1}{\eta} e^{-\|\mathbf{q} - \mathbf{p}_i\|^2 / \eta^2}$ is the weight defined in (6). A close inspection of (10) suggests that this choice of $\boldsymbol{\mu}$ is the non-local mean solution using the best k patches, and $\boldsymbol{\Sigma}$ is the covariance of those patches.

The advantage of the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ defined in (10) is that they are 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*

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

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

Therefore, by solving the group sparsity problem (5), we not only obtain \mathbf{U} , but also obtain $\boldsymbol{\Lambda}$, automatically. The overall algorithm is given in Algorithm 1.

Algorithm 1 Proposed Algorithm

Input: noisy patch \mathbf{q} , and similar patches $\mathbf{p}_1, \dots, \mathbf{p}_k$

Output: estimate $\hat{\mathbf{p}}$

Learn \mathbf{U} and $\boldsymbol{\Lambda}$

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

Denoise: $\hat{\mathbf{p}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{q}$

It is instructive to compare the proposed Bayesian approach with existing methods. First, we note that many PCA-based patch denoising algorithms [2, 3, 5, 16] can be generalized under our Bayesian

framework. In those methods, the prior $f(\mathbf{p})$ is a delta function centered at the initial guess $\bar{\mathbf{p}}$: $f(\mathbf{p}) = \delta(\mathbf{p} - \bar{\mathbf{p}})$, whereas in our method, the prior $f(\mathbf{p})$ has finite covariances.

As a comparison to methods using generic databases such as [4, 9, 21, 26], 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 a spatially adaptive prior $f(\mathbf{p})$ for different patches. As a result, 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 several 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 so that patches can be searched over multiple images. As for NLM, instead of using all patches in the database, we consider only the best k patches following [27]. For EPLL, we consider 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 (which are single image denoising algorithms) 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 (size 161×145) using a collection of 9 arbitrarily chosen text images containing texts of the same font sizes. The purpose of the experiment is to simulate the case where we want to denoise a noisy document with the help of other similar but non-identical texts.

Fig. 1 shows the results of the experiment. We observe that among all the methods, our proposed method yields the highest PSNR and SSIM values. Our PSNR is 6dB better than the benchmark BM3D (internal) algorithm.

3.3. Denoising Multiview Images

The second experiment considers the scenario of capturing images using a multiview camera system where one of the views is corrupted. We add i.i.d. Gaussian noise to one of the five multiview images, and use the rest of the images to denoise.

Fig. 2 illustrates the denoising results of the ‘‘Barn’’ and ‘‘Cone’’ multiview datasets, which indicate that our proposed method yields much better PSNR than BM3D. In Fig. 3 we plot and compare the PSNR values over a range of noise levels. The proposed method is consistently better than its counterparts.

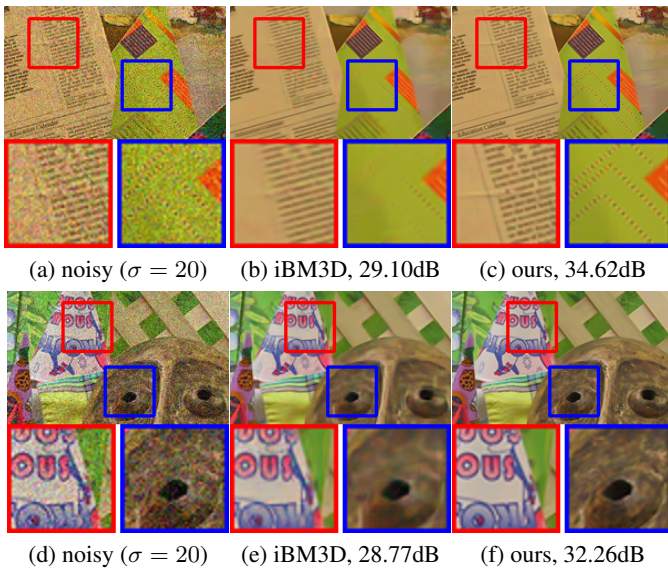


Fig. 2: Multiview image denoising using the proposed method and internal BM3D. [Top] ‘‘Barn’’; [Bottom] ‘‘Cone’’.

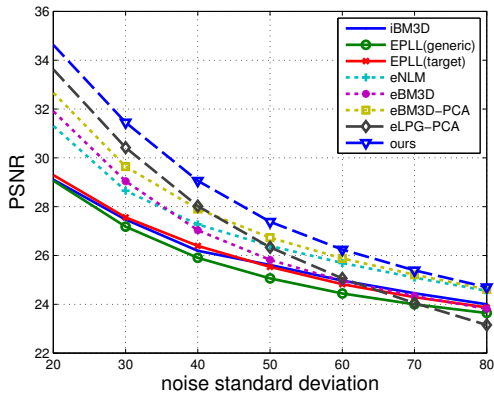


Fig. 3: Multiview image denoising for ‘‘Barn’’: PSNR vs noise levels. In this plot, each PSNR value is averaged over 8 independent trials to reduce the bias due to a particular noise realization.

3.4. Denoising Human Faces

The third experiment considers denoising face images using a dataset from [28]. We simulate the denoising task by adding noise to a randomly chosen image and use other images in the database for denoising. The faces are not pre-processed, and so 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 denoised results are shown in Fig. 4, and the PSNR curves are shown in Fig. 5.



Fig. 4: Face denoising of Gore dataset [28]. [Top] Database images; [Bottom] Denoising results.

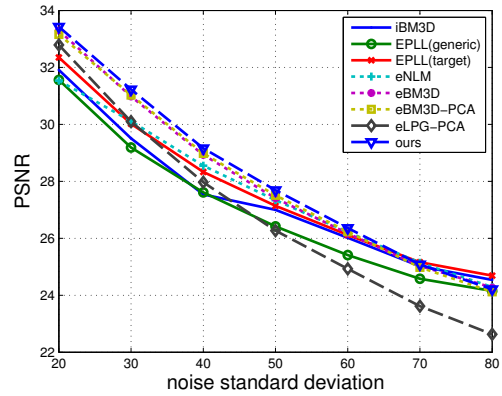


Fig. 5: Face denoising results: PSNR vs noise level. In this plot, each PSNR value is averaged over 8 independent trials to reduce the bias due to a particular noise realization.

4. CONCLUSION

Classical image denoising methods based on single noisy input and 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, only similar 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, pp. 1603–1607.
- [12] S.H. Chan, T. Zickler, and Y.M. Lu, "Monte Carlo non-local means: Random sampling for large-scale image filtering," submitted, arXiv:1312.7366, 2013.
- [13] 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.
- [14] I. Ram, M. Elad, and I. Cohen, "Image processing using smooth ordering of its patches," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2764–2774, July 2013.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] S. Roth and M.J. Black, "Fields of experts," *Intl. J. Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.
- [20] L. Sun and J. Hays, "Super-resolution from internet-scale scene matching," in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, 2012.
- [21] 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.
- [22] P. Milanfar, "A tour of modern image filtering," *IEEE Signal Processing Magazine*, vol. 30, pp. 106–128, Jan. 2013.
- [23] H. Talebi, X. Zhu, and P. Milanfar, "How to SAIF-ly boost denoising performance," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1470–1485, April 2013.
- [24] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, Feb 2014.
- [25] P. Milanfar, "Symmetrizing smoothing filters," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 263–284, 2013.
- [26] 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.
- [27] 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.
- [28] 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.